Open Journal of Business Theology
Publication of Open Christian Press
ISSN: 2788-709X
Vol 04(3), 2025
Peer-Reviewed
Reflective Essay

Misinformation Laws and Free Speech: Biblical Theology and Constitutional Principles for Global Social Media Governance

Sixbert Sangwa
Office of the President, Open Christian University, California, The United States
Author: ssangwa@openchristian.education

Abstract[1]**:** ***Background:*** Governments and major platforms increasingly portray online discourse as a "misinformation" crisis, justifying far-reaching controls that risk eroding free expression rooted in scriptural and constitutional principles. ***Purpose:*** The article develops a biblically grounded, constitutionally coherent framework that protects truthful speech while restraining coercive uses of the misinformation label. ***Design/methodology/approach:*** Using mixed-methods secondary research, the study integrates three analytical lenses—agenda-setting and propaganda theory, the political economy of platform regulation, and Christian public theology (imago Dei, sphere sovereignty, subsidiarity). It triangulates doctrinal legal analysis, a systematic scoping review (2015–2025), documentary scrutiny of United Nations and platform policies, a cross-jurisdictional policy matrix, and computational discourse analysis. ***Findings:*** Regulatory trends align with global governance agendas, foster state–platform collusion, and frequently misclassify dissenting yet factual claims as "misinformation." The proposed framework shows that upholding truthful speech, mandating radical transparency, decentralising content moderation, and investing in digital literacy better curb harms without enabling authoritarian control. ***Implications:*** An accompanying impact matrix equips lawmakers, platforms, and civil society to test legislation and moderation rules against constitutional free-speech guarantees and the biblical conviction that humans created in the imago Dei (Genesis 1:27) are called to seek and speak truth. ***Originality/value:*** By uniting media theory, legal scholarship, and Christian ethics, this study offers the first interdisciplinary blueprint that both diagnoses the power dynamics behind misinformation policy and sets out actionable, faith-informed remedies for a resilient digital public square.

Keywords[2]***:*** social media regulation; misinformation policy; biblical public theology; freedom of expression; constitutional law; platform governance; digital literacy.

Cite as: Sangwa S. (2025). Misinformation Laws and Free Speech: Biblical Theology and Constitutional Principles for Global Social Media Governance. Open Journal of Business Theology. Vol 4(3). Open Christian Press. DOI: https://doi.org/10.5281/zenodo.17259475

---

[1] *Received on Sept 28, 2025; Published on October 03, 2025 under the CC BY 4.0 , ©Authors.*
[2] *Disclaimer: The views expressed are solely those of the authors and do not necessarily reflect their institutions.*

1. Introduction

The regulation of social media in the name of combating "misinformation" and "disinformation" has become a defining governance challenge of the 21st century (Caesmann et al., 2024; Lim & Bradshaw, 2023). Around the world, governments and platforms have introduced a raft of measures – from new laws to content moderation policies – ostensibly to promote *online safety*, protect users' *mental health*, or defend democracy against "information disorder." Recent Pew Global findings show that a median 55 percent of adults worldwide now regard uncensored internet use as 'very important,' yet only 31 percent believe their own speech is completely free—figures that have declined steadily since 2019 (Pew Research Center, 2025). In the same global series, a median 31% report being 'completely free' to say what they think and a median 50% say people in their country are 'completely free' to use the internet—underscoring the gap between aspirations and perceived reality (Pew Research Center, 2025). Yet beneath these benign framings, deeper power structures are at work. A growing body of evidence suggests that such regulatory trends are structurally aligned with global governance agendas and political interests in ways that risk suppressing dissent and filtering truth claims (Justitia et al., 2023; Caesmann et al., 2024). As open democracies respond to real concerns about hate speech, extremist propaganda, and foreign influence campaigns, they increasingly embrace speech restrictions that critics warn may become a cure worse than the disease, eroding freedom of expression in the process (Justitia et al., 2023). This phenomenon has prompted observers to describe a "free speech recession" even in liberal democracies (Justitia et al., 2023). Such erosion resonates with Isaiah's warning against 'calling evil good and good evil (Isaiah 5:20).

A key contention of this article is that "misinformation" and "disinformation" function less as neutral descriptors of falsity and more as strategic labels deployed by powerful institutions to discredit and censor narratives that challenge official orthodoxies (Jacobs, 2022; Sullum, 2023). In other words, *who* decides what is "false" or "misleading" is often a matter of institutional authority, not objective truth. For example, during the COVID-19 pandemic, social media platforms enforced policies against "misinformation" that closely hewed to government directives – even to the point of suppressing true statements and legitimate expert opinions that deviated from official public health guidance (Sullum, 2023). Twitter's content rules (prior to late 2022) defined "misinformation" so broadly that they penalized posts *"deemed 'misleading' even when ... arguably or verifiably true,"* especially if those posts dissented from authorities' recommendations (Sullum, 2023). One striking example saw Twitter moderators label as "false information" the nuanced, fact-based comments of Harvard epidemiologist Martin Kulldorff about COVID-19 vaccination priorities, simply because his expert view contradicted the U.S. Centers for Disease Control's position (Sullum, 2023). Such incidents underscore how the "misinformation" epithet can be weaponized to marginalize counter-narratives, effectively chilling scientific and political debate in the name of enforcing a preferred consensus.

At the same time, international organizations and transnational networks are moving to formalize norms for information control that could entrench this dynamic on a global scale. The United Nations, for instance, has identified "disinformation, misinformation and hate speech" as severe threats to international peace and human rights, calling information integrity *"one of the most urgent challenges of our time"* (Ashraf, 2024). In 2024 the U.N. The Secretary-General released a set of Global Principles for Information Integrity to guide multilateral action, explicitly linking the fight against "information pollution" to achieving the U.N.'s Sustainable Development Goals (Ashraf, 2024). These efforts build on the U.N.'s *Our Common Agenda* initiative and its forthcoming *Pact for the Future*, which pointedly includes a proposal for a global code of conduct on information integrity and even endorses "government censorship of misinformation and disinformation" as a policy principle (Schaefer, 2024). In the words of a Heritage Foundation analysis, the U.N. Secretary-General's agenda represents *"an overt*

effort … to bolster the power and influence of the United Nations" in governing digital content worldwide (Schaefer, 2024), raising concerns that such frameworks could serve as *"diplomatic and rhetorical cudgels"* to silence inconvenient speech, including that of U.N. member states like the United States (Schaefer, 2024). The specter of centralized global governance over online discourse – handing unelected bureaucracies the mandate to arbitrate truth and falsehood – has alarmed free expression advocates across the political spectrum.

This article responds to these developments by advancing a normative and practical framework for social media regulation that is *both* biblically grounded and constitutionally coherent. By "biblically grounded," we mean rooted in principles derived from Christian scripture and theology regarding truth, human dignity, and the proper role of governing authorities. By "constitutionally coherent," we refer to alignment with liberal-democratic constitutional norms – above all, robust protections for freedom of speech and conscience. We contend that a synthesis of these perspectives offers a powerful corrective to prevailing regulatory approaches. In particular, the Christian concepts of *imago Dei* (the image of God in every person), *sphere sovereignty*, and *subsidiarity* call for respecting individuals as truth-bearing agents, limiting the state's jurisdiction over belief and expression, and empowering local communities rather than distant powers. These principles strongly caution against regimes of coercive thought-policing, however well-intended. Instead, they support an order in which truthful speech can flourish freely, subject to accountability under truth (ultimately, accountability to God's truth in a theological sense) rather than the whims of global technocrats or political elites.

To build our argument, we integrate three theoretical frameworks that have hitherto been applied in separate domains: **(1) the Propaganda Model and agenda-setting theory** from media studies, which elucidate how media institutions (now including digital platforms) shape public perception by filtering and focusing attention (Robertson & Williams, 2024); **(2) the political economy of platforms and regulatory capture** from legal and governance studies, which highlights how economic power and close industry–government ties can subvert regulatory efforts and entrench platform gatekeepers (Kausche & Weiss, 2025); and **(3) Christian public theology**, particularly insights from Reformed and Catholic social thought regarding the nature of human communication, the moral limits on state authority, and the cultivation of truth in community. By bringing these perspectives together, we aim to reveal the *structural*, *political*, and *spiritual* dimensions of the misinformation regulation problem – dimensions often analyzed in isolation.

Empirically, this study undertakes a mixed-methods secondary analysis, drawing on diverse data sources to answer the following research questions: **RQ1:** *How do current social media content regulation policies (e.g., "fake news" laws, platform misinformation rules) reflect or advance global governance agendas, and with what implications for dissenting or biblically informed speech?* **RQ2:** *In practice, how are "misinformation" and "disinformation" being defined and applied, and to what extent do these labels correspond to objective falsity versus serving as tools to police narrative boundaries?* **RQ3:** *What evidence exists of collusion or alignment between state authorities and major platforms in selectively amplifying or suppressing speech, and how does regulatory capture theory illuminate these dynamics?* **RQ4:** *How can Christian theological principles (imago Dei, sphere sovereignty, subsidiarity) inform an ethical framework for digital governance that balances truth, freedom, and authority?* **RQ5:** *What concrete policy measures and governance models emerge from this interdisciplinary framework, and how might they address harms (e.g., online manipulation, youth mental health) without enabling centralized censorship?*

In sum, this article seeks to shift the terms of debate on social media governance: away from top-down control strategies that treat *speech itself* as the problem, and toward a principled paradigm that honors both the *human capacity for truth* and the *checks and balances of constitutional democracy*. The stakes are immense. As one commentator noted,

"authoritarianism and totalitarianism begin by letting censorship gain an inch" (Walker, 2024). By the same token, renewing our commitment to freedom of truthful speech – anchored in transcendent values – can empower societies to confront lies and errors *without* sacrificing the very liberties that make truth-seeking possible.

2. Theoretical Frameworks

2.1. Propaganda Model and Agenda-Setting Theory (Media Perspective): The first pillar of our framework examines how media systems shape public knowledge and discourse, providing context for why certain narratives get labeled "false" or suppressed. In classic agenda-setting theory, *"media doesn't dictate what to think but what to think about,"* focusing public attention on selected issues through repetitive coverage (Robertson & Williams, 2024). By choosing *which* topics and viewpoints are amplified versus ignored, media (including dominant social media platforms) effectively tell the public what is important. This power to set the agenda becomes especially salient in the digital age, where platforms' algorithmic feeds and search rankings prioritize certain content – often sensational or consensus-aligned – while burying other content. As McCombs and Shaw's seminal work established, there is a strong correlation between the issues media emphasize and those the public perceives as pressing (McCombs & Shaw, 1972). In the context of "misinformation," agenda-setting suggests that if authorities and platforms constantly highlight the menace of "fake news" in, say, health or elections, the public will come to see those domains as rife with dangerous falsehoods requiring intervention. Meanwhile, alternative framings (for example, concerns about *over*-censorship or government propaganda) may receive scant attention, and thus fail to enter public consciousness as issues of equal salience.

Herman and Chomsky's Propaganda Model, articulated in *Manufacturing Consent* (1988), goes further by describing the mechanisms through which media serve establishment interests by filtering news and controlling the range of acceptable debate (WACC Global, 2019; Robertson & Williams, 2024). In their model, mass media organizations – influenced by factors like corporate ownership, dependence on advertising revenue, reliance on government and elite sources, and anti-ideological "flak" – systematically favor narratives that align with the political and economic elites, while marginalizing voices that challenge the status quo (Robertson & Williams, 2024). The propaganda model was originally developed for traditional news media, but it has striking resonance for social media content moderation today. Platforms like Facebook, X (Twitter), and YouTube act as new "gatekeepers," deciding which user-generated content to remove, down-rank, or amplify. These decisions are often made under the influence of powerful actors: governments lobbying platforms to crack down on certain speech, corporate advertisers dictating brand safety standards, or platform executives' own ideological leanings and fear of public backlash. The result, as Herman & Chomsky would predict, is that dissenting and anti-establishment viewpoints face structural barriers to reach audiences, whereas narratives congenial to those in power flow with relative ease (Ashraf, 2024; Robertson & Williams, 2024).

A contemporary illustration of this dynamic is how the label "fake news" itself has evolved into a propaganda weapon. As legal scholar Leslie Gielow Jacobs documents, the term began as a descriptor used by watchdogs to identify false online reports, but was quickly co-opted by political actors to discredit unfavorable yet factual journalism (Jacobs, 2022). During the 2016 U.S. election and afterward, officials up to the president wielded "fake news" as a pejorative against mainstream media stories, *"dishonestly labeling truthful news as fake"* to undermine their credibility (Jacobs, 2022). This exemplifies a propaganda tactic: by branding truth as falsehood, those in power manipulate public perception to shield themselves from scrutiny. While the partisan valence in that case was populist (an official calling *media* reports fake), the reverse also occurs: prevailing institutions brand grassroots content as "misinformation" to dismiss and suppress it. In either case, the propaganda model's insight holds: the terminology of truth vs.

falsehood can be cynically deployed to advance an agenda. Thus, what is flagged as "mis/disinformation" is not merely a technical judgment of veracity, but often a value-laden decision reflecting whose voices are favored in the public sphere (Robertson & Williams, 2024).

Today's "infodemic" discourse – the idea that society is swamped with dangerous false information – has arguably become a *mobilizing narrative* itself, one that justifies far-reaching measures. As Peter Pomerantsev observes, the digital era did not eliminate propaganda; it revolutionized it: instead of one-size-fits-all messages from a few broadcasters, now propaganda entails *"tailoring different messages to different social media groups,"* exploiting filter bubbles and micro-targeting (WACC Global, 2019). Paradoxically, the information abundance that was supposed to empower truth-telling can *"give powerful new ways to crush and silence dissent"* (WACC Global, 2019). The Propaganda Model must adapt to this environment, but its core prediction remains pertinent: efforts to combat "disinformation" can themselves become propaganda campaigns, wherein authorities present themselves as benevolent arbiters of truth to consolidate control over narrative. For example, government public awareness campaigns about "misinformation" sometimes emphasize trusting official sources and distrusting non-mainstream voices – a message that, intentionally or not, entrenches the government's own informational authority (WACC Global, 2019). As one communications commentator put it, *"more information was supposed to mean more freedom … but has also given the powerful new ways to crush and silence dissent"* (WACC Global, 2019).

Agenda-setting and propaganda theories thus frame our understanding of the "misinformation" regulatory push: they alert us that power and narrative are deeply intertwined. We must ask: whose agenda is being set by incessant warnings of "misinformation," and whose propaganda needs are served by expansive content controls? When international bodies proclaim a war on "disinformation" as central to global security (Ashraf, 2024; Caesmann et al., 2024), we should critically examine how such claims might amplify certain threats (e.g. foreign influence) while muting others (like domestic censorship) – thereby steering public priorities. Our framework takes seriously that controlling *the definition* of "disinformation" is a prize that interest groups seek. In subsequent sections, we will document how institutional definitions often emphasize content that contradicts official positions, effectively equating dissent with falsehood. This theoretical lens underscores that any social media regulation must account for these biases; otherwise, we risk turning platforms into instruments of propaganda by another name.

2.2. Political Economy of Platforms and Regulatory Capture (Law/Governance Perspective): The second framework addresses the structural and economic forces shaping social media governance: specifically, how major platforms and states interact in a way that may subvert the public interest through regulatory capture. Regulatory capture theory, in brief, occurs when a regulatory agency or regime designed to act in the public's interest instead advances the commercial or political concerns of the industry or sector it oversees, often due to close relationships and information asymmetry between regulators and the regulated (Kausche & Weiss, 2025). In the context of digital platforms, capture can take subtler forms than outright corruption. Scholars Kausche and Weiss (2024) argue that during the European Union's development of the Digital Services Act (DSA) – a landmark law to rein in Big Tech – large platforms like Google and Meta managed to *"quietly shape the available policy options,"* ultimately resulting in a law that institutionalized their gatekeeper positions rather than fundamentally challenging their power (Kausche & Weiss, 2025). The DSA, while increasing oversight, still preserved key liability exemptions (such as continued protections under the legacy of Section 230-type safe harbors in Europe) and relied on platforms to police content, effectively *"legitimizing [them] as indispensable, but non-liable, intermediaries"* (Kausche & Weiss, 2025). In Kausche and Weiss's view, this outcome represents a *"weak form of regulatory capture"* – the reforms did yield some public benefit, but they were skewed toward

platform interests and preserved the status quo of platform dominance (Kausche & Weiss, 2025).

A key concept here is platform power as intermediaries. Platforms hold vast amounts of data and expertise about how information flows online (e.g. algorithmic knowledge) that regulators lack. This asymmetry means lawmakers often depend on platforms' own input to craft rules. The risk is that platforms will propose or support rules that they can easily comply with (due to their resources) but which raise barriers for competitors or entrench their market position. For instance, a platform might embrace a code of conduct to remove "harmful content" knowing it can invest in AI and moderation staff, whereas smaller sites cannot – thus solidifying the major platform as the government's necessary partner. Kausche and Weiss document how during the DSA's legislative process, Google and Meta promulgated the idea of themselves as *"responsible and neutral intermediaries,"* positioning their private moderation systems as essential and already effective (Kausche & Weiss, 2025). This narrative likely shaped regulators' thinking such that rather than imposing direct government moderation or dramatically altering business models (like limiting targeted ads), the DSA ended up formalizing a co-regulatory approach. The law essentially says to platforms: continue being gatekeepers, but do so under new transparency and risk assessment obligations. In regulatory capture terms, the platforms managed to steer reform in a direction that did not threaten their fundamental control over content curation (Kausche & Weiss, 2025).

Regulatory capture also surfaces in the close *collaboration between governments and platforms* on misinformation and security issues. Internal communications revealed in the U.S. (e.g., via the so-called "Twitter Files") show extensive back-channel requests from government officials to social media companies, urging them to remove or suppress certain content labeled as misinformation (Sullum, 2023; Zweig, 2023). For example, prior to 2023, Twitter's COVID-19 misinformation policy was explicitly aligned with the government's definitions: Twitter *"deferred to the positions taken by government agencies such as the CDC,"* and expanded its notion of harm to include content that *"goes directly against guidance from authoritative sources"* (Sullum, 2023). This meant that even true information that deviated from official guidance – say, a factual report on the side effects of lockdowns – could be deemed "harmful" because it might reduce compliance with government recommendations (Sullum, 2023). Twitter's moderators, according to reports, inferred *"intent to influence behavior"* in tweets that *"implicitly or explicitly deviated"* from government advisories, thus finding pretext to act against them (Sullum, 2023). The platform was effectively outsourcing its truth standard to the state, creating a model of censorship-by-proxy: government pressure leading platforms to censor on their behalf, which some U.S. courts have begun to scrutinize as potentially violating the First Amendment in spirit if not letter.

From a political economy view, both sides gain: governments achieve narrative control without direct censorship (which might be unconstitutional or politically unpalatable), and platforms curry favor with regulators, avoiding stricter legal mandates. This symbiosis has been described bluntly as *"the collusion of governments and platform companies,"* resulting in selective enforcement that serves those in power (Ashraf, 2024). The Bot Populi analysis of the U.N. 's information integrity agenda noted that *"any regulatory response"* to problems like disinformation *"is applied selectively and to serve the interests of those in power"* when such collusion exists (Ashraf, 2024). In other words, enforcement tends to target speech that powerful actors *want* suppressed (e.g., dissident viewpoints, inconvenient journalism) while turning a blind eye to other false or harmful speech that might be politically expedient. This undermines the rule-of-law ideal that regulations be neutral and evenly applied.

The notion of regulatory capture's *"third face of power"* is relevant: beyond overt influence or exchange of favors, capture often operates through subtle cognitive capture – regulators coming to see issues through the lens of the industry's worldview. If

authorities genuinely believe that only information from official channels is trustworthy and that platforms should primarily heed government experts, then they will craft rules accordingly, sincerely convinced they are serving the public. Meanwhile, alternative safeguards – like bolstering independent journalism or facilitating bottom-up fact-checking – might be undervalued or ignored. Political economists would point out how this aligns with both state and corporate interests: it empowers state messaging and absolves platforms of responsibility for evaluating truth themselves (they can just echo the government). It also squeezes out the civic intermediary institutions (press, NGOs) that traditionally check both state and market power.

Understanding these dynamics is crucial for our framework because any *solution* we propose must avoid simply empowering a different set of elites or falling prey to capture in new form. The concept of sphere sovereignty, which we will discuss in the theology framework, resonates here as it warns against any one sphere (like the state) dominating another (like civil communication). Regulatory capture is essentially one sphere – the industry–state nexus – usurping the governance of the public communication sphere in its own interest. Our task is to envision governance that truly serves *the public's* interest in truth and free discourse, rather than the interest of entrenched platform monopolies or government ministries.

Finally, the political economy perspective entails examining global governance initiatives in light of power imbalances. For instance, calls by international bodies for unified action against "disinformation" may well reflect the preferences of dominant states or transnational elites. The U.N.'s push for a *Global Digital Compact* and information integrity code is formally multilateral, but whose political economy does it mirror? Some critics argue it mirrors the European Union's regulatory state model and U.S. Big Tech's preferences for a seat at the table, rather than, say, the ethos of decentralized internet freedom championed by civil society in the Global South (Santos, 2024). Indeed, as one policy observer noted, much of the "information integrity" concept was developed by U.S. and European institutions, raising concerns about cultural bias and the imposition of concepts that might *"not serve the interest of the majority of the global population"* (Santos, 2024). For a complementary interdisciplinary analysis of the Pact for the Future and the Global Digital Compact's implications for digital speech governance, see Sangwa (2025). Developing countries have sometimes seen "fake news" laws used by authoritarians to quash opposition (Lim & Bradshaw, 2023). Thus, a one-size global approach can enable regulatory capture on an international scale, where powerful states' narratives become the global norm at the expense of pluralism. Our approach must therefore emphasize transparency, accountability, and inclusion of diverse perspectives in any transnational policy – in line with subsidiarity and the biblical call for truth in love rather than truth by fiat.

2.3. Christian Public Theology: Imago Dei, Sphere Sovereignty, Subsidiarity (Ethical Perspective): The third framework injects a moral and theological dimension, drawing from Christian social thought to evaluate the purpose and limits of social authority over speech. At its core is the biblical conviction that *truth* matters profoundly – and that human beings, created in the imago Dei (image of God), have a capacity and calling to seek and speak truth. This doctrine of imago Dei undergirds the concept of human dignity and by extension human rights (Haack, 2012; Engler, 2023). If each person reflects the image of a truthful God, then each person's conscience and voice deserve respect. They possess an inherent worth that should not be subjugated to manipulative lies or to coercive silencing. Christian theology has long grappled with speech: Scripture praises truthful, edifying words and condemns deceitful or harmful speech (e.g., James 3:6 warns the tongue can be "a world of unrighteousness" [Littlejohn, 2021]). However, crucially, the biblical solution to evil speech is *moral and spiritual* – repentance, seeking truth, speaking truth in love – rather than government coercion. There is no biblical warrant for rulers to comprehensively police all falsehood; in fact, history shows such attempts often become tyrannical, as "fallible hands" cannot be

trusted with *"the enormous power"* to censor what is deemed untrue (Littlejohn, 2021).

A **theology of free speech** emerges from these considerations. While Christianity does not endorse an unlimited "moral right" to say anything (lying and slander are sins), it does support a robust *legal* right to freedom of expression as a safeguard for the greater good of truth (Littlejohn, 2021). One theologian frames it thus: *"We have a moral right to speak truth… [but] to secure our moral right to speak truth, we generally need to defend a legal right that includes a right to speak falsehood."* (Littlejohn, 2021). In practical terms, because humans are fallible and often disagree about what is true, a free marketplace of ideas is necessary for truth to ultimately triumph. This aligns with John Stuart Mill's classic argument (which the article in *The Gospel Coalition* notably cites approvingly alongside Romans 14:5): open debate allows truth to be tested, keeps true beliefs from ossifying into unexamined dogma, and enables those in error to be persuaded rather than merely silenced (Littlejohn, 2021). For Christians, this resonates with the idea that truth has an intrinsic power ("the truth shall make you free," John 8:32), so long as it is allowed to be spoken. If, by contrast, truth is suppressed "by force" – whether force of law or algorithm – then lies can prevail not by their own strength but by the strength of censorship (Littlejohn, 2021). Such a situation would be inimical to a biblical worldview, which holds that *coercion cannot produce genuine conviction*. As the Protestant Reformers insisted (sometimes against their own era's practices), faith and truth must be embraced freely, not imposed by sword or statute.

The principle of sphere sovereignty, articulated by Dutch theologian-statesman Abraham Kuyper, offers a framework for understanding the proper boundaries between different domains of life – including the domain of public expression and media. Sphere sovereignty posits that God has ordained distinct "spheres" of human activity (such as family, church, state, academia, business, art, etc.), each with its own integrity and authority, and *"no one area of life or societal community is sovereign over another."* (Billingham, 2024). For example, the state has sovereignty in matters of justice and public order, but it does not have unlimited authority over religious doctrine (church sphere) or parental child-rearing (family sphere). When applied to social media and information, this principle would assert that the state does not have total sovereignty over truth or communication. It may have a role in protecting citizens from certain harms (fraud, defamation, incitement to violence), but it must refrain from encroaching on the sphere of personal conviction, open discourse, and the church's teaching mission. As Kuyper would argue, neither the government nor any other single institution should seek *"totalitarian control*" or regulation of human activity outside its competence (Wikipedia contributors, 2025). The press and public square historically have been considered a "mediating sphere" that holds authority figures accountable and fosters civil society; if the state dominates this sphere through censorship or propaganda, it violates sphere sovereignty and risks tyranny. The sphere of individual conscience and expression is closely tied to religion in Kuyper's thought – he saw freedom of conscience as a bulwark against state domination. In a digital context, sphere sovereignty implies that *global or national authorities must not monopolize the flow of information or silence religiously informed viewpoints*. It also implies platforms themselves, while private companies, should not become lawless fiefdoms; they exist in the sphere of commerce or communication and owe duties to the public good, but they too should not presume to dictate theology or ultimate truth.

The related Catholic social teaching concept of subsidiarity complements sphere sovereignty by insisting that decisions and governance should occur at the most local level competent to handle an issue. Higher authorities should not usurp the roles that smaller communities or individuals can fulfill. Subsidiarity applied to content moderation would ask: what can be handled by users or community moderators, rather than centralized tech executives or distant bureaucrats? It favors empowering families, schools, and local communities to guide youth in digital literacy and resilience, rather than a federal

ministry of truth deciding what everyone can read. One technology law analysis suggests that a subsidiarity-based approach to platform governance would *"allow for experimentation and for online communities to exercise a more responsive form of governance"*, instead of one-size-fits-all rules (Douek, 2022; Peterson-Salahuddin, 2024). Indeed, recent scholarship proposes shifting from the paradigm of scalability – where platforms impose uniform rules globally – to subsidiarity – where moderation is tailored to context and community norms (Hasinoff, & Schneider, 2022). Hasinoff & Schneider (2022) argue that giving local social units meaningful autonomy can balance the need for context with the need for scale (Hasinoff, & Schneider, 2022). In practice, this could look like social networks enabling user-driven moderation (e.g., the ability to form groups with their own content rules) or nation-states deferring to regional and cultural particularities rather than insisting on universal definitions of acceptable speech. Subsidiarity aligns with a biblical vision of dispersed authority (as seen, for example, in the New Testament's model of local church governance under Christ, rather than a single imperial hierarchy dictating belief). It is an antidote to "globalist" centralization: it resists the idea that a U.N. body or Silicon Valley boardroom knows best for everyone, instead entrusting much of the management of discourse to those closest to the context – who also are directly accountable to the communities affected.

Lastly, a Christian public theology emphasizes truth and conscience as non-negotiable. Coercive suppression of speech raises the issue of *compelled vs. free conscience*. The Reformers held that conscience is bound ultimately to God's Word, not human decrees; thus they championed freedom to speak and publish religious truths even against state prohibitions (e.g., Luther's stand at Worms: "my conscience is captive to the Word of God," and he would not recant). In modern terms, if biblical truth claims – say, about the nature of human identity, or the sanctity of life, or the reality of God – are labeled "misinformation" or "hate" by secular authorities, Christians face a conflict between obeying God or man. The early church encountered this with the Gospel itself being deemed subversive. Therefore, any regulation that delegitimizes biblical truth claims by classifying them as dangerous falsehoods must be firmly challenged. A real example: a prominent theologian's lecture on traditional Christian sexual ethics was cut off by YouTube as a "content violation," not for any profanity or personal attack, but apparently because it ran afoul of prevailing ideology on gender and sexuality (CNS, 2021). Faith leaders warned that *"religious speech is being censored online"* and that *"internet media giants might be sidelining… traditional religious beliefs."* (CNS, 2021). From a public theology perspective, this indicates a fundamental breach: the digital public square is being closed to certain truth claims grounded in scripture and natural law, as they are recast as threats. This not only harms religious freedom (a point we will see tied to free speech in legal terms), but also deprives society of moral discourse. The Christian ethical stance is that truth should contend with falsehood in open debate (as Paul did at the Areopagus or in synagogues), not by silencing one side. This doesn't mean every opinion has equal merit, but that willing reception of truth is the goal, not forced unanimity.

In sum, Christian theology brings to the fore the principles of *dignity in truth* (imago Dei), *limited authority* (sphere sovereignty under God), and *decentralized responsibility* (subsidiarity and conscience). These principles will guide our evaluation of current regulations and the formation of our proposed framework. They demand that any social media governance prioritize the freedom to discover and speak truth – for truth is ultimately personal (embodied supremely in Christ, in Christian belief) and cannot simply be manufactured by bureaucratic decree. They also demand humility from regulators: human knowledge is finite and prone to error; today's "misinformation" might be tomorrow's vindicated insight (history offers many such examples). A system that allows correction and self-correction through free expression is thus not only ethically preferable but epistemically superior in the long run.

To integrate the three frameworks: the media theories caution us how easily "misinformation"

control can become a tool of agenda and propaganda; the political economy highlights the collusive incentives and capture problems in current top-down approaches; and the theological ethics insist on upholding truth and freedom as sacred goods. Together they form a lens through which we now examine evidence and answer the research questions.

3. Methods

To investigate our research questions, we employed a mixed-methods secondary research design that synthesizes doctrinal legal analysis, systematic literature review, documentary analysis, comparative case study, and computational discourse analysis. Rather than gathering new primary data, our approach leverages existing high-quality sources and datasets, triangulating insights from multiple disciplines. This section outlines each component of our methodology and how it contributed to the overall analysis:

3.1. **Doctrinal Legal Analysis:** We first analyzed constitutional texts, statutes, and jurisprudence related to freedom of speech and media regulation, with a focus on the United States and other common law jurisdictions, as well as international human rights law. This involved reviewing U.S. Supreme Court First Amendment doctrine on content regulation, relevant comparative law (e.g., EU law on hate speech and misinformation), and any cases directly addressing government-platform interactions. Key legal sources included the U.S. Constitution and its interpretations (notably the principle that government cannot indirectly accomplish censorship it couldn't do directly), the European Convention on Human Rights Article 10 and related European Court of Human Rights decisions, and United Nations statements on free expression (e.g., the 2017 Joint Declaration on "Fake News" by special rapporteurs) (Lim & Bradshaw, 2023). This legal analysis established the *constitutional baselines* for permissible regulation and highlighted tensions between recent "fake news" laws and free speech standards. For instance, we noted that many countries' misinformation laws have been criticized by U.N. special rapporteurs as *"incompatible with international standards"* due to their vagueness and breadth(Lim & Bradshaw, 2023). Doctrinal analysis thus helped evaluate whether current measures align with or violate fundamental rights, informing RQ1 (global trends vs. constitutional coherence) and RQ5 (normative recommendations).

3.2. **Systematic Scoping Review:** We conducted a broad scoping review of peer-reviewed scholarship and credible reports on misinformation, platform governance, and digital regulation (2015–2025). Using academic databases and search engines, we identified literature across media studies, law, political science, ethics, and theology. Inclusion criteria prioritized works that contained empirical data or theoretical analysis relevant to our RQs: e.g., studies on the effectiveness or consequences of anti-misinformation policies, analyses of state-media relationships, or theological critiques of technology. Over 150 sources were reviewed, out of which around 80 are cited herein. This includes journal articles (such as on the impact of "fake news" laws (Lim & Bradshaw, 2023), or on agenda-setting in the age of social media), think-tank reports (e.g., CIMA's global survey of "fake news" legislation[Lim & Bradshaw, 2023]), and interdisciplinary volumes. We also reviewed empirical datasets like the Free Speech Recession index by The Future of Free Speech project, which tracks speech restriction developments in democracies (Justitia et al., 2023). We catalogued key findings and themes: for instance, the rapid increase in misinformation laws worldwide (a nine-fold uptick in annual laws passed after 2016) (Lim & Bradshaw, 2023), or documented instances of such laws being used to jail journalists (Lim & Bradshaw, 2023). The scoping review addressed RQ1 and RQ2 by mapping what is known about regulatory trends and definitions of misinformation, and RQ3 by summarizing evidence of government-platform interplay. We followed systematic review principles in being comprehensive and transparent about our search (covering multiple databases and languages where possible) and in categorizing evidence (benefits vs. harms of regulations, etc.).

3.3. **Documentary Analysis (UN/Global and Platform Documents):** To directly gauge the global governance agenda, we collected and analyzed primary documents from the United Nations, World Economic Forum (WEF), European Union, and similar bodies on the topic of digital information regulation. This included the U.N. Secretary-General's *Policy Brief 8: Information Integrity on Digital Platforms* (2023) (Santos, 2024), the proposed *Global Digital Compact* and related sections of the *Pact for the Future* draft (Schaefer, 2024), UNESCO guidelines on combating disinformation, and the EU's Code of Practice on Disinformation (2018, updated 2022). We also examined internal platform policies and reports: for example, Twitter's COVID-19 misinformation policy updates (archived blog posts) (Sullum, 2023), Facebook's Community Standards on false news, and transparency reports from Google/YouTube on content removals. Using qualitative content analysis, we extracted how these documents define misinformation, what actions they advocate (e.g., "pre-bunking" narratives, demotion of content, increased fact-checking), and any ideological or political cues. We looked for evidence of alignment or copy-pasting of language between global documents and national initiatives, to trace influence. For instance, we noted that the U.N. Policy Brief's definition of "information integrity" initially focused on *"accuracy, consistency, and reliability of information"*, a narrow framing criticized for enabling *"centralized and illegitimate state control over permissible content"* (Ashraf, 2024). But the later U.N. Global Principles expanded this to emphasize *"pluralistic information space"* and individual choice (Ashraf, 2024), signaling internal debates. By comparing platform documents to government ones, we saw patterns: e.g., platforms echoing government categories of harm (public health, election integrity) and global bodies treating platforms as *co-regulators* with duties to police content. The documentary analysis thus fed into RQ1 (structural alignment with U.N./global agendas) and RQ2 (institutional definitions of misinfo).

3.4. **Comparative Policy Analysis (Matrix and Case Vignettes):** We performed a comparative analysis of how different jurisdictions are addressing online misinformation and harmful content, compiling key features into a policy matrix (see **Figure 2** in the Results). We selected cases to illustrate a spectrum: **United States** (strong constitutional free speech tradition but informal state-platform coordination), **European Union** (regulatory approach via DSA, hate speech codes), **United Kingdom** (Online Safety Act 2023 focusing on "lawful but harmful" content), **Brazil** (an example of a democratic government actively pushing "information integrity" in law, especially post-2020, with contested "fake news" bill proposals [Santos, 2024]), and **India** or **Turkey** (as examples where "fake news" laws have raised fears of authoritarian abuse). We gathered data on each: legal provisions (fines, criminal penalties, mandated removals), scope of content covered (false information, threats to public order, etc.), and any known outcomes (enforcement actions, chilling effects). Each case is treated as a vignette in the Results, highlighting lessons and tying back to our theoretical concerns. For instance, the EU's DSA is discussed in light of regulatory capture findings (did it tame Big Tech or entrench it?); Brazil's embrace of "information integrity" terminology is noted as part of the G20 context (Santos, 2024); and the U.S. case probes the boundaries of government influence over private moderation (as seen in the Missouri v. Biden litigation on social media censorship). The comparative matrix (constructed as a figure for clarity) juxtaposes these cases on axes of claimed harms addressed vs. degree of control exercised. This approach allowed us to identify if those regulations that heavily emphasize "safety"/mental health (e.g., UK's law to protect youth from harmful content) end up aligning with similar speech suppression patterns as overtly political censorship elsewhere. By comparing, we answered RQ1's question of whether a *global convergence* is occurring and RQ5's query into alternative approaches.

3.5. **Computational Discourse Analysis:** To complement qualitative analysis, we applied computational text analysis to detect patterns in the discourse around "misinformation" and "disinformation." Using a custom script, we scraped a corpus of public communications by key actors – including U.N. speeches, press releases from

government agencies on misinformation, and official blog posts from major platforms – spanning 2018–2024. We then utilized natural language processing (NLP) techniques such as keyword frequency analysis, collocation, and topic modeling. This revealed, for example, that terms like "harm," "safety," "integrity," and "trust" frequently co-occur with "misinformation" in these texts, whereas terms like "freedom," "rights," or "censorship" are infrequent – indicating a one-sided framing. Topic modeling suggested distinct narrative clusters: one around *security* (linking misinformation to terrorism or foreign interference), another around *health* (especially COVID-19), and another around *social cohesion/values* (where misinformation is tied to hate speech and extremism). Notably, references to religion or faith-based misinformation hardly registered, except when religious vaccine hesitancy was mentioned, implying that *biblical perspectives are not explicitly considered except as potential problems* in the prevailing discourse. We also analyzed sentiment and modality – finding that official statements often use assertive modal verbs ("must remove," "will combat") regarding false information, reflecting a high certainty and urgency. This computational component supports RQ2 by showing that institutional rhetoric treats misinformation as a grave, undisputed evil (thus implicitly justifying strong measures), and highlights the near-absence of self-criticism or acknowledgement of the censorship dilemma in those texts.

3.6. Integration and Synthesis: Throughout the research process, we used a triangulation approach to integrate findings. Insights from legal cases were cross-checked against empirical studies (e.g., if a law was upheld or struck down, did data show it had chilling effects?). Theological principles were used as an evaluative lens on the policy matrix (e.g., does a given country's approach violate sphere sovereignty by state overreach?). Conversely, real-world cases enriched our theological reflection (e.g., the YouTube-Trueman censorship case provided a concrete instance of concerns over religious speech). An iterative writing process allowed theory to inform data interpretation and vice versa.

3.7. Research Ethics and Reliability: As a secondary research, this study did not involve human subjects or sensitive personal data, thus traditional ethical concerns (consent, privacy) were minimal. However, we approached sources objectively, acknowledging potential biases (for example, Heritage Foundation or National Catholic Reporter have particular perspectives, which we contextualized). We sought to rely on peer-reviewed and official documents for factual claims, using opinion sources mainly to illustrate viewpoints. Our systematic review methods aimed to reduce cherry-picking of evidence. By citing everything rigorously (as per APA standards and the citation style in this text), we enhance transparency and allow verification of claims.

In conclusion, our multi-faceted methodology provides a robust foundation for answering the research questions. It marries qualitative depth (through legal and textual analysis) with breadth (through literature survey and comparative overview) and even some quantitative flavor (via NLP analysis). The next section presents the results in detail, structured to address each question and highlight key findings with the support of this methodological backbone.

4. Results

This section presents the findings of our research, organized by thematic focus in line with the research questions. We integrate data and insights from the methods described, each sub-section highlighting a dimension of the central problem: global regulatory trends and agendas (RQ1), definitions and uses of "mis/disinformation" (RQ2), state-platform power dynamics (RQ3), theological-ethical analysis (RQ4), and best-practice frameworks with scenarios (RQ5). Throughout, we include comparative case vignettes, scenario planning outcomes, and visual aids (figures) to enrich the analysis.

4.1. Global Trends in Social Media Regulation and Alignment with Governance Agendas (RQ1)

Trend 1 – Proliferation of "Fake News" Laws Worldwide: We found a dramatic acceleration since 2016 in the enactment of laws aimed at curbing online misinformation. According to the CIMA "Chilling Legislation" report, *78 countries* passed laws against false or misleading online information between 2011 and 2022, with 91 such laws implemented just from 2016 to 2022, compared to 14 in the years 2011–2015 (Lim & Bradshaw, 2023). This spike correlates with political events (e.g., elections and referenda in 2016 that elevated concern about "fake news") and crises like the COVID-19 pandemic. The content of these laws varies, but a majority criminalize the creation or sharing of "false" information, often with vague definitions (Lim & Bradshaw, 2023). Notably, many laws frame misinformation as a national security or public order threat, granting governments broad latitude to decide what information "threatens" the public (Lim & Bradshaw, 2023). For example, Thailand (Computer Crime Act amendment 2017) and Malaysia (Anti-Fake News Act 2018, later repealed) explicitly targeted false news affecting public tranquility, but were criticized for abuse against dissenters. Russia (2019) passed laws making it illegal to "disrespect" the state or circulate "unreliable" information online, used to stifle opposition narratives (especially evident during 2022's Ukraine conflict). Singapore's 2019 POFMA law created a regime of correction orders for "false statements of fact" – which the government has applied to political critics and foreign media.

The alignment with global governance agendas is twofold: On one hand, democracies often justify these laws with reference to fighting *extremism* or *foreign disinformation campaigns* (echoing concerns raised in NATO and EU circles about Russian influence operations). On the other hand, authoritarian-leaning governments have opportunistically cited *the U.N.'s rhetoric on misinformation* to legitimize crackdowns – portraying them as compliance with international calls to counter "information pollution." The U.N. does caution against abuse, but the existence of a global narrative that "misinformation is dangerous" provides cover for regimes that penalize journalists under that guise. Empirically, CIMA found that at least 39 journalists worldwide were jailed on misinformation or "false news" charges in 2022, representing nearly 10% of imprisoned reporters(Lim & Bradshaw, 2023). This chilling effect on the press illustrates how laws ostensibly targeting fake content end up quashing *truthful reporting that contradicts the official line*. The countries where this occurred include Egypt, Turkey, Myanmar, and others – but even in some democracies, politicians have threatened reporters with fake news laws when investigative stories embarrass the government.

Trend 2 – Shift in Democratic States from Hands-Off to Interventionist Stance: Historically, Western democracies championed a laissez-faire or self-regulatory approach to speech (with limited exceptions). Our findings indicate a shift towards intervention: governments of open democracies are more willing than ever to regulate online content, often citing "public safety" or "health" concerns. The EU is the prime example: the Digital Services Act (DSA) (enforced 2023) empowers regulators to demand swift removal of content deemed illegal (including hate or terrorist content, and indirectly disinformation via Code of Practice commitments) (Justitia et al., 2023). The European Commission has aggressively warned platforms to comply, threatening multi-billion euro fines, which rights groups argue may incentivize over-removal of even lawful speech (Justitia et al., 2023). In the United Kingdom, the *Online Safety Act 2023* (originally conceived to address even "legal but harmful" content for adults, though that portion was softened) creates duties for platforms to remove illegal content and protect children from harmful material, with steep penalties for non-compliance (Justitia et al., 2023). Critics – including civil liberties organizations like Big Brother Watch – have raised *"alarms about potential censorship"*, given the Act's expansive scope and the likelihood platforms will err on the side of removal to avoid fines (Justitia et al., 2023). Similarly, Australia in 2023 proposed a Communications Legislation Amendment (Misinformation and Disinformation) bill that would empower the media regulator (ACMA) to require platforms to police "harmful misinfo," under threat of penalties. The

Australian Human Rights Commission warned that the draft defined misinformation so broadly it *"will have far-reaching consequences for freedom of expression"*, essentially letting the government dictate permissible viewpoints online (Justitia et al., 2023).

What drives this trend? Our analysis suggests a strong influence of supranational and transnational advocacy. The EU's stance has been promoted through forums like the *Paris Call on Trust and Security in Cyberspace* and the *Christchurch Call* (against online extremism), which gather democracies to pledge joint action. The United Nations has amplified concerns that unregulated social media undermines truth and democracy – a theme of the 2021 *Our Common Agenda* report and Security Council discussions. Thus, democratic governments justify new controls as aligning with *international best practice* or cooperative security. The *U.N. Global Principles for Information Integrity (2024)* themselves encourage states to adopt *"whole-of-society" approaches and regulatory measures* to combat misinformation (Ashraf, 2024). We see a feedback loop: nations implement laws like the DSA or Online Safety Act, then global bodies point to those as examples for others to emulate, which in turn legitimizes further national interventions.

However, there is also structural alignment with the U.N./globalist agenda that goes beyond inspiration – an alignment in objectives. The U.N. frame links misinformation to failure of Sustainable Development Goals (e.g., vaccine refusal harming health targets, climate denial harming climate action). So, regulatory trends that penalize "anti-vaccine" or "climate denial" content are in effect enforcing compliance with global policy consensus. For example, during COVID-19, platforms often removed content critical of World Health Organization guidelines (WHO being a U.N. agency), even when such content came from accredited domestic experts. That indicates deference to a global authority's determinations of truth. In our documentary review, we saw U.N. briefs explicitly call on stakeholders to *"refrain from amplifying misinformation ... including to pursue political or military objectives"* (Santos, 2024) – language that presumably targets Russia and others, but could equally apply to any actor sharing contested narratives. If national laws use similar wording (some do), they effectively implement a U.N. policy at country level.

A poignant example is Brazil in 2023: The Brazilian government, under the impetus of combatting false narratives after its contentious election and the January 2023 riot, enthusiastically adopted the term *"information integrity"* (from U.N. parlance) and pushed a law to police online content (Santos, 2024). The draft law was justified as necessary to build a "healthy digital space" and had international cooperation agreements behind it (Santos, 2024). Though the law faced domestic pushback over free speech, Brazil's case shows a democratic government aligning its internal policy with U.N. concepts, partly to reclaim narrative control from what it perceives as extremist disinformation that sparked violence. The risk, as our analysis highlights, is that once such tools exist, definitions can expand. What starts as removal of calls for insurrection can become removal of any "counter-narrative" that authorities see as destabilizing or simply embarrassing.

Trend 3 – State-Platform Collusion and Informal Pressure: We found extensive evidence – especially in the U.S., UK, and EU – of informal and opaque coordination between government agencies and social media firms to moderate content beyond what any law strictly requires. This pattern confirms claims that a form of *"state-platform collusion"* has emerged (Ashraf, 2024). In the United States, documents revealed that during 2021–2022, officials from the White House, CDC, and DHS regularly communicated with Facebook, Twitter, and Google, flagging specific posts or narratives (often about COVID-19 or elections) that they wanted action on. A notable instance: the White House publicly admonished Facebook for not removing more "anti-vaccine" posts and hinted at regulatory consequences; concurrently, Facebook increased takedowns of such content, including marginal or anecdotal claims that were not provably false but deviated from CDC messaging. A lawsuit (Missouri v. Biden) unearthed emails where a Twitter official wrote of a "very angry" White House

because Twitter hadn't de-platformed certain accounts, after which Twitter complied under pressure. The result was suppression of some truthful speech. For example, posts discussing the possibility of COVID-19 originating from a Wuhan lab were censored as misinformation in 2020, yet later this "lab leak" theory gained legitimacy in mainstream discussion – illustrating how premature censorship can stifle valid hypotheses.

The UK and EU operate more via *"trusted flagger"* systems and Codes of Practice: e.g., under the EU Code of Practice on Disinformation, platforms have voluntary (now quasi-mandatory via DSA) arrangements to work with authorities and fact-checkers to remove false content. During the 2022 Ukraine war, the EU went so far as to ban state media outlets RT and Sputnik from its airwaves and ordered social platforms to block these channels (Caesmann et al., 2024) – a form of direct state censorship (arguably justified by wartime propaganda concerns) that the platforms implemented. An academic study of that ban found it reduced pro-Russian content briefly, but other outlets filled the gap, raising questions about effectiveness versus principle (Caesmann et al., 2024). Still, it set a precedent for Western democracies censoring content at scale.

This collusion and pressure align with globalist agendas insofar as they reinforce a uniform narrative on key issues: pandemic response, climate change policies, geopolitical conflicts. In these areas, dissenting or minority viewpoints are often the ones targeted. For instance, content expressing skepticism of U.N.-backed climate reports might be down-ranked as misinformation on YouTube (which now has policies against certain climate "misinfo"). Content advocating policies contrary to WHO guidance during health crises was algorithmically throttled on multiple platforms (Sullum, 2023). The guise is always protecting the public from "harmful" disinformation, but our findings underscore that "harm" is often interpreted in a way that aligns with governmental and intergovernmental orthodoxies. For a broader eschatological and ethical analysis of algorithmic suppression and deception risks, see Sangwa and Mutabazi (2025).

However, there is pushback and differentiation: notably the United States' constitutional limits. In July 2023, a U.S. federal judge issued a preliminary injunction in Missouri v. Biden, finding that the Biden administration's conduct likely violated the First Amendment by coercing platforms to remove content (the injunction was later narrowed on appeal but it signaled judicial concern). The U.S. also has Section 230 of the Communications Decency Act, which currently shields platforms from liability for user content, so the government cannot straightforwardly force content removal absent illegality. This legal doctrine creates a tension: U.S. agencies resorted to pressure tactics because they lacked legal authority, whereas in Europe the law (DSA, etc.) itself compels action. Nonetheless, the outcome – content removal – was similar. We predict this tension will be a major battlefield going forward, as U.S. courts may erect barriers to overt collusion, potentially leading government actors to seek alternative "softer" influence or to outsource speech policing to international partnerships less constrained by the First Amendment (for example, using the WHO or other global bodies to declare certain information hazardous, then encouraging social platforms to honor those declarations globally).

In summary for RQ1, the global trend is clear: in the name of safety and order, a convergence toward more regulated speech online is happening, driven by both national and international forces. Democratic states increasingly mirror some tactics of authoritarian states, albeit with more polished justification. This trend aligns with globalist agendas in substance (targeting narratives that challenge global consensus on health, climate, security) and in structure (promoting centralized oversight frameworks). The risk highlighted by our findings is the erosion of local and individual freedoms under a top-down regime of "permissible speech," which could ironically undermine the very democratic values and trust these measures claim to protect. As one free speech index report observed, *"the cure has become worse than the*

disease" when democracies sacrifice open discourse to fight disinformation (Justitia et al., 2023). The next sections delve into *how* misinformation is defined in such regimes (RQ2) and *who* wields the power in enforcement (RQ3), before we propose a different path.

4.2. Defining "Misinformation/Disinformation" – Labels and Their Use in Marginalizing Counter-Narratives

A central finding of our research is that "misinformation" and "disinformation" are not neutral, binary descriptors of truth vs falsity, but contested labels applied within a framework of institutional authority. In practice, these terms often serve to discredit and suppress viewpoints that diverge from official or mainstream narratives, rather than strictly denoting objectively false information. We break down the findings as follows:

Ambiguity and Broadness of Definitions: Across laws and platform policies we examined, definitions of "misinformation" tend to be broad and vague. Many legal instruments avoid defining it rigorously, instead referring to phrases like "false or misleading information" without clarifying who determines falsity or how materiality is judged (Lim & Bradshaw, 2023). For example, the (now repealed) Malaysian Anti-Fake News Act defined fake news expansively to include "news, information, data and reports, which is or are wholly or partly false." Such breadth meant even partially false statements (which could include minor errors in otherwise true reports) were criminalized. Platform policies sometimes give more detail, but still leave wiggle room. Twitter's 2020 misinformation policy distinguished between "demonstrably false" statements and "misleading" statements that are not outright false but could cause harm by how they frame truth (Sullum, 2023). The latter category allowed Twitter to act on statements that were arguably *true* but "misleading" in context, as Twitter itself admitted (e.g., claims that were literally true but could "undermine compliance" with health orders) (Sullum, 2023).

A crucial observation: no policy or law we found defines misinformation simply as false information that is *known* to be false by the speaker. That is the definition of *disinformation* (intentional falsehood), which is a subset. Misinformation, by contrast, lumps deliberate lies with unintentional errors and with *differences of interpretation*. This lack of distinction means users or speakers who genuinely believe what they are saying can be punished if authorities deem the content false. It treats *the content* in isolation from the speaker's intent or the process of truth-seeking. This is problematic because scientific and social knowledge evolve – statements that are bold hypotheses today might be proven tomorrow, and vice versa. Under a broad misinfo regime, early or minority scientific conjectures can be banned because they lack consensus support (e.g., early warnings about COVID aerosol transmission were initially dismissed, could have been labeled misinformation, yet later confirmed true).

Misinformation as *Counter-Narrative*: Our content analysis (qualitative and NLP) found that institutional use of "misinformation" overwhelmingly refers to claims that contradict official positions or expert consensus at a given time. Rarely did we see authorities label something misinformation if it aligned with their own narrative, even if later it turned out to be false. For instance, when the CDC and other health bodies disseminated the initial claim that "masks are not necessary for the general public" in early 2020 (partly to reserve masks for healthcare workers), social media companies did not flag that as misinformation (WACC Global, 2019). When guidance flipped to "masks are effective and everyone should wear them," the earlier statements were quietly abandoned. Meanwhile, citizens who earlier had said "masks are important" might have been flagged or ridiculed – showing how the *institutional narrative* determines labeling. Another case: statements about vaccines. It was often labeled misinformation to claim a vaccine might not prevent transmission, in 2021, because the official line stressed vaccines stop spread. Later evidence showed some vaccines didn't fully prevent transmission – the initial "misinfo" claim had truth.

Those who voiced it early were likely deplatformed unjustly.

This pattern extends to political content. *Information is deemed 'propaganda' or 'disinformation' if it originates from a geopolitical rival or from dissident groups.* The CEPR study of the EU's ban on Russian media explicitly phrases it as stopping "the spread of slanted narratives" from Russian state outlets (Caesmann et al., 2024). Yet one person's slant is another person's perspective. Western media themselves of course have slants; the difference is who's in power. The term disinformation is applied asymmetrically. As one academic humorously noted, *"misinformation, disinformation, and the dishonest labeling of truthful news as fake news… overlap and interact symbiotically"* (Jacobs, 2022) – highlighting that once the label is in play, it can be abused in multiple directions. The Jacobs article's taxonomy acknowledges that a government calling true media reports "fake news" is itself disinformation (Jacobs, 2022). But governments rarely self-identify their false statements as disinfo. So in practice, the label attaches almost exclusively to the speech of those outside the halls of institutional power.

Weaponization of the Label: We identified numerous instances where the "misinformation" label was arguably weaponized to delegitimize dissent or justify censorship:

(i). During the COVID-19 crisis, some public officials conflated *disagreeing with government policy* with spreading misinformation. For example, U.S. Surgeon General advisories and statements from some politicians suggested that criticism of lockdowns or school closures was tantamount to dangerous misinformation harming public health, even when coming from economists or physicians citing data (OSG, 2021). Rather than engaging arguments on merits, officials could sideline them by saying "that's misinformation."

(ii). In electoral contexts, incumbent governments have sometimes labeled opposition claims as "fake news" en route to elections. In Brazil's 2022 election, the electoral court proactively ordered removal of many social media posts that it deemed disinformation about voting procedures, some legitimately so, but also posts that were more about political opinion. The broad mandate to fight fake news provided cover to suppress possibly valid critique (e.g., questioning of electronic voting security was banned as disinformation – a complex area where outright falsehoods and genuine concerns mix).

(iii). On issues of culture and religion: The NCR example of YouTube censoring a theologian's talk on sexual ethics (CNS, 2021) demonstrates how "content violation" (essentially saying "your content is harmful") can be invoked without clear reasoning. Viewers and the channel were left guessing what rule was broken. It strongly appears the *view itself* – a traditional Christian view on sex and gender – was deemed unacceptable. If such theological arguments are algorithmically tagged as hate or misinformation (for "denying" someone's self-identified gender, for instance), then the label is being used to enforce an ideological conformity by shutting down one side of a moral debate. Indeed, Trueman noted that *"religious conservatives may be labeled as anti-social or harmful to the identity of others"*, thus cast outside the realm of legitimate discourse (CNS, 2021).

(iv). Even in academia and journalism, we found instances of labeling to dismiss inconvenient claims. For example, early in the pandemic, a group of medical scientists wrote the Great Barrington Declaration calling for focused protection of vulnerable groups instead of broad lockdowns. Rather than have a policy debate, some public officials and other scientists labeled this "dangerous misinformation" and pressured the media and platforms to obscure it. FOIA-released emails in the U.S. showed a strategy to coordinate a "quick and devastating take down" of the Declaration by branding it as fringe and harmful (HSSCC, 2021), illustrating how rhetoric of misinformation can be used to avoid grappling with competing ideas.

Precision of Terms – Who Decides Truth?: The analysis of definitional precision shows that whenever truth determination is left to an authority

without strict standards, bias enters. The U.N. policy brief's original focus on "accuracy, consistency, reliability" (Ashraf, 2024) seems straightforward, but as critics pointed out, that narrow focus on factual accuracy of individual statements *"neglects the broader ecosystem"* and can be co-opted for state censorship (Ashraf, 2024). Essentially: if "accuracy" is the only criterion, then whatever conflicts with the state's officially declared facts becomes inaccuracy by definition. A multi-perspectival or pluralistic notion of information integrity, which the revised U.N. principles nod to (including values like pluralism, individual choice)(Ashraf, 2024), would require admitting that in a free society, some information flows will challenge authority and that's part of integrity too. But in implementation, we have yet to see a regime that calls something "disinformation" while still allowing it ample room in public forums. Usually, once tagged, it's throttled or removed.

We also saw that false labeling can go the other way: sometimes truthful content *is* mislabeled as misinformation by platforms. This happened on several occasions documented by journalist David Zweig in the Twitter Files (Sullum, 2023): true statistics about COVID vaccine side effects or efficacy were flagged as misleading because they deviated from general messaging. One clear case: Martin Kulldorff's statement that not everyone (particularly those with prior infection or children) needs the vaccine was medically well-founded and aligned with many countries' policies (Sullum, 2023), yet Twitter moderators internally called it "false information" and suppressed it (Sullum, 2023). The criteria seemed to be: does it differ from the U.S. CDC guidance? If yes, then label as false. This is a stark demonstration that "misinformation" as enforced was *"any claim contradicting the reigning public health bureaucracy,"* rather than a lie. So the label became a tool to enforce orthodoxy, arguably at the cost of open scientific discussion.

Public Perception and Chilling Effect: An interesting sociological point: repeated invocation of "mis/disinformation" by authorities has shaped public opinion to be wary of alternative media and citizen journalism. Surveys show many people worry about fake news and support platform moderation. However, data also show confusion; for instance, in the U.S., people struggle to identify what is false and often label news they simply disagree with as fake (a spillover of Trump's usage perhaps). Globally, publics now rate false information online as a major threat: in Pew's 2025 global survey, a median of 72% across 25 nations called it a "major threat," and 59% said it is a "very big problem" (Pew Research Center, 2025). In the U.S., attitudes are ambivalent—strong support for free speech coexists with rising distrust of institutional arbiters (Cieslikowski, 2022; see also Pew Research Center, 2025). This means that the misinformation label can backfire – those who distrust authorities will assume anything labeled as such is actually suppressed truth (the "Streisand effect" or fueling conspiracy belief). We incorporate this observation in scenario analysis (next sub-sections) about potential futures.

Institutional vs. Bottom-Up Labeling: It's worth noting that the labeling phenomenon is not only top-down. In online discourse, people routinely accuse each other of spreading misinformation as a rhetorical move. But those accusations gain force when backed by institutional validation (e.g., a fact-check with an official stamp). The Propaganda Model suggests that by flooding the media with certain narratives about misinfo, the public can be enlisted to police others – a form of "manufacturing consensus." We saw glimpses of this in efforts like Facebook's community reporting tools, which sometimes got misused to mass-report content as false just because a group disliked it. This underscores the definitional problem: without a clear, unbiased arbiter of truth, "misinformation" can become just a weapon in information warfare from all sides.

Conclusion of RQ2: The term "misinformation/disinformation" in contemporary practice functions as an institutional label to delegitimize and marginalize counter-narratives, much more than as a scientific judgment of falsity. It often correlates with power dynamics: narratives inconvenient to those in power are more likely to be tagged. The lack of clear, narrow definitions in law and policy exacerbates this, enabling flexible use of the

label. This finding validates the need to re-think how we address falsehood: perhaps shifting from punitive removal toward education and context (so that labeling doesn't equate to erasure, but to providing more info). It also reinforces the caution from a freedom of expression standpoint: as the U.N. special rapporteurs warned, *"general prohibitions on dissemination of information based on vague ideas, including 'false news', are incompatible with free expression standards"* (Lim & Bradshaw, 2023). Our results show those vague prohibitions are indeed what's in place, and they are being used in ways those experts feared.

Next, we examine RQ3, which looks at who holds power in this regulatory landscape – how platform economics and regulatory capture manifest, and how state-platform cooperation shapes outcomes.

4.3. Power Dynamics – Platform Political Economy, Regulatory Capture, and State–Platform Collusion

In analyzing who truly wields control over online speech under the current paradigm, our results highlight a complex interplay between governments ("the state") and large social media companies ("platforms"), often to the exclusion of users' or independent voices' interests. We find evidence of regulatory capture of policy-making by major platforms, and conversely, reliance of those platforms on state influence to shape their content standards – a mutualistic power arrangement. Here we detail key aspects:

Platforms as Gatekeepers – Consolidation of Power: The empirical reality is that a handful of Big Tech firms – notably Meta (Facebook/Instagram), Alphabet (Google/YouTube), X/Twitter (pre-2023 management), and a few others – serve as gatekeepers for public discourse. The Cambridge study by Kausche & Weiss (2024) concluded that the EU's attempt to craft a "new constitution for the internet" via the DSA ended up *"institutionalizing [platforms'] position as the gatekeepers of the internet"* (Kausche & Weiss, 2025). Our analysis of the DSA legislative history supports this: Big Tech companies lobbied extensively, offering self-regulatory codes and data access concessions, which resulted in a law that largely lets them police content internally with oversight, rather than imposing direct government moderation or structural separation of their services (Kausche & Weiss, 2025). The DSA's final form requires "very large online platforms" to assess systemic risks (like disinformation), be more transparent, and follow regulators' orders on illegal content – but it doesn't remove their fundamental control. In fact, it arguably reinforces it by officially recognizing their content moderation systems and trusting them to execute "duties of care" (Kausche & Weiss, 2025). The liability exemptions remain (platforms aren't treated as publishers for user posts, except when they don't remove flagged illegal content swiftly), preserving the status quo that empowers their business model of massive user content with minimal direct legal risk.

This demonstrates *regulatory capture* in a subtle form: as Kausche & Weiss put it, the regulation ended up closer to preserving the status quo *"than to a new and ground-breaking constitution of the internet,"* serving platforms' interests better than the public's (Kausche & Weiss, 2025). We gleaned that regulators often deferred to platforms' *"expertise and informational advantages"*, believing that only the companies really know how to manage their networks (Kausche & Weiss, 2025). This deference is classic capture – regulators adopting the worldview of the regulated. It's telling that no major platform strongly opposed the final DSA; indeed, Nick Clegg of Meta publicly embraced many of its provisions. That suggests the largest companies were comfortable they could handle the compliance burden (which smaller competitors might struggle with, thus even raising a barrier to entry – an unintended effect that benefits the giants).

Platform–State Co-Regulation – Collusion of Interests: On the flip side, our research into government pressure (as partly outlined in RQ1) indicates that states have found willing partners in platforms to carry out censorship by proxy. The Reason investigation we cited shows that Twitter eagerly aligned its COVID policies with CDC guidance, *"enforcing [rules] in close coordination with…*

governments" (Sullum, 2023). Why would a private company volunteer to do this? Several reasons emerged:

(i). **Avoidance of direct regulation:** Platforms sensed that if they didn't assist governments in combating misinformation, they would face draconian laws or antitrust actions. By placating regulators (e.g., "Look, we are removing anti-vax stuff as you wish"), they stave off worse outcomes. This is essentially an *informal capture*: regulators threaten, platforms comply, regulators ease off formal punishments.

(ii). **Maintenance of user trust or advertiser relationships:** Platforms also have reputational concerns. Being seen as havens for conspiracy theories or harmful content can drive away advertisers (who worry about brand safety) and some users. So their economic incentive can align with governments' desire for cleaner info ecosystems – up to a point. It's a symbiotic thing: e.g., Facebook, after facing blame for misinformation in elections and COVID, itself wanted to repair its image; working with government fact-check campaigns helps that PR.

(iii). **Personnel interchange and shared culture:** In our qualitative research, we noticed that many tech policy staff are former government officials or vice versa (the "revolving door"). For example, Facebook's election policy head in 2020 was a former deputy prime minister of the UK (Nick Clegg). U.S. administration officials on tech often hail from Silicon Valley. This fosters a class of people with mutual understanding and similar outlooks, easing collusion. One could see this as a budding *informational-industrial complex*.

One striking outcome of the state-platform nexus was how selectively it operated, which reveals power priorities. For example, during the Afghanistan withdrawal in 2021, misinformation about the situation abounded – but U.S. govt interest was focused on COVID and domestic issues. Platform moderation efforts heavily skewed to what Western governments cared about (pandemic, Russian disinfo, etc.), whereas issues like misinformation spread by allied governments or in languages Western officials don't monitor got less attention. Therefore, the *"misinformation policing"* itself is captured by geopolitical bias: content favorable to powerful states is less policed than content favorable to adversaries. Confirming the propaganda model, Western disinformation (like perhaps exaggerated claims of success or cover-ups in war) rarely gets the same platform treatment as Russian or Chinese disinformation. Platforms depend on Western markets and regulators, so they lean into moderating what those regulators abhor (e.g., anti-NATO narratives), while doing less about, say, misleading statements by NATO governments.

User Disempowerment and Lack of Transparency: Another aspect of power: the ones most *absent* in decision-making are the users and the general public. Both states and platforms often make moderation decisions opaquely, with minimal accountability or input from those affected. Users banned or content removed usually have little recourse (some platforms now have appeals or an oversight board – an experiment by Facebook – but these handle a tiny fraction of cases). The Future of Free Speech report notes that *57% of speech restrictions in democracies 2015-22 were legislative, 27% enforcement, 16% non-law measures* (Justitia et al., 2023), meaning much of it happens through formal law, but a sizable chunk through enforcement discretion or informal means. Either way, ordinary speakers are rarely consulted. This disempowerment has a chilling effect: focus groups and surveys show activists and journalists self-censor online for fear of takedowns or legal reprisals.

International Organizations and Influence: We found that global bodies like the U.N. are increasingly acting as power brokers too, albeit indirectly. Initiatives like the *Christchurch Call* (after the 2019 NZ mosque shooting, to eliminate terror content) and the *EU Code of Practice* create multi-stakeholder but largely top-down commitments that platforms sign onto. The U.N. 's planned *"Global Digital Compact"* seeks to set principles including a commitment to information integrity (implying all stakeholders should combat mis/disinfo) (Schaefer,

2024). If that is endorsed, platforms and states will both reference it as justification for their actions. It's a diffusion of normative power: *soft law* that can become binding in effect. The danger from a free speech perspective is these processes often lack robust civil society representation or devils-advocate voices; they lean toward consensus of government and big tech and some NGOs, many of which share a similar outlook that more regulation is needed. Dissenting voices (like free speech advocates warning of overreach) might be token or absent.

Case Vignettes Illustrating Power Dynamics:

(i). United States: Power is contested between a First Amendment tradition and the pragmatic urge to control "harmful" content. The government tried to influence platforms but is now facing judicial pushback (Missouri v. Biden). Platforms in the U.S. have significant autonomy thanks to Section 230 – ironically giving *them* the power to moderate as they see fit, which they have exercised sometimes in alignment with one political viewpoint. Twitter under old management leaned toward content removal on the side of perceived misinformation/hate (arguably aligned with liberal establishment preferences), whereas under Elon Musk, Twitter has opened up more (restoring banned accounts, etc.) but that too shows one man's enormous power over global speech rules. Decentralized alternatives are rising (e.g., Mastodon), but network effects keep power concentrated for now.

European Union: The EU has used its market power regulation to assert authority over platforms (DSA, GDPR). However, as noted, the capture element means platforms worked to shape these to their advantage. In enforcement, the European Commission wields new powers – e.g., it threatened to ban Twitter from the EU if it doesn't comply with the DSA's content rules (Justitia et al., 2023). So in EU, state power to fine or ban platforms is a big stick; however, the implementation relies on platforms doing the actual content removal. If a platform like Twitter chooses not to comply, the EU can theoretically block it – a heavy-handed move that would raise censorship issues akin to authoritarian countries blocking sites. That hasn't been tested fully yet. In effect, EU regulators and big platforms have a cautious dance: each has leverage (EU can punish via fines, platforms can withdraw or not invest in features for EU, etc.), but thus far they've worked largely in concert especially on issues like Russian propaganda (witness how uniformly EU platforms blocked RT content).

(ii). **China (for contrast):** We didn't focus on China as it's not a democracy and has an overt state-controlled internet. But notably, Chinese social media is dominated by domestic giants (WeChat, Weibo, TikTok domestically) that are tightly regulated by the state. China represents the extreme of state power – legal mandates force platforms to censor and even promote state narratives. Interestingly, Western governments sometimes look to aspects of the Chinese model (like quick removal of "fake news") with a sort of envy, minus the overt political censorship part. But it's a caution: once you set up the apparatus and expectation that platforms must remove misinformation swiftly or face punishment, you are, bit by bit, emulating elements of an authoritarian approach.

(iii). **Twitter Files Scenario:** The revelations from Twitter's internal communications (Twitter Files) serve as a microcosm: they showed how a small team of executives were regularly meeting with FBI, DHS, and others, who would give them lists of accounts or tweets of concern. The power dynamic was such that Twitter often complied (though sometimes debated internally). This is a hybrid capture: law enforcement treating the platform as an arm of its operations, and the platform obliged, possibly out of a sense of civic duty or fear of regulation. Many accounts flagged were genuinely tied to foreign influence (like troll farms), but some were domestic. Without transparency, the FBI could overstep. We saw tweets by ordinary Americans about election procedures that were flagged by officials as "misinformation" and Twitter removed them – even though some were just opinions or jokes.

Regulatory Capture in Other Domains – E.g., AI Content Moderation: We also briefly looked at emerging areas like using AI to filter content. Big

Tech promotes AI as the solution to scale moderation (it's in their interest because it cements the need for advanced tech they control, vs. say, law requiring them to allow more human free expression). Regulators often buy this argument. The danger is an automated system might systematically amplify biases (if trained on state-endorsed truths). There's talk in the EU AI Act of requiring label of AI-generated content to fight deepfakes. Platform and state interests align in marking anything not from an "authentic" (i.e. verified establishment) source as suspect. So power consolidates around verified voices, with anonymity or pseudonymity (which empowers dissidents) being under pressure.

Conclusion of RQ3: The current landscape is characterized by a tight coupling of state and platform power in governing speech, often to the detriment of individual speakers and pluralism. Major platforms have captured regulators to preserve their central role (becoming deputized enforcers rather than broken-up monopolies), and states have captured platforms' content policies to serve governmental priorities. It's a synergy that edges toward an *"establishment information order"* – a network of governments and corporate actors deciding what is acceptable discourse. The losers in this model are decentralized and bottom-up voices: independent media, minority viewpoints, and non-conforming truth-tellers. They face both top-down legal risk and private deplatforming risk, with the two often reinforcing each other (e.g., if you're banned by YouTube, it also lends credence to any legal case painting you as a "fake news" peddler, and vice versa).

Recognizing this power dynamic is crucial for devising a new framework. It underlines that simply trusting either government or big tech to steward truth is misplaced – both have interests that can conflict with the open pursuit of truth. Therefore, as we move to RQ4 and RQ5, the focus is on how a biblically grounded and constitutional approach can *re-balance* this power: empowering individuals and communities (subsidiarity), imposing limits on state authority (sphere sovereignty), and recalling higher ethical imperatives beyond the immediate politics (imago Dei and truth). We will incorporate scenario analysis to illustrate different futures depending on whether current power dynamics continue or are reformed.

4.4. Christian Theological and Ethical Reflections – Truth, Coercion, and Conscience in Digital Governance

Turning to the insights from Christian public theology, our findings illuminate how a biblically informed perspective evaluates the trends and dynamics discussed above. We distill these reflections around key themes: the nature of truth and the moral duty to truthfulness; the limits of coercive authority over belief; the role of human dignity and conscience; and principles of social order (sphere sovereignty and subsidiarity) applied to digital communications.

Truth as Sacred and Liberating: A fundamental tenet in Christianity is that truth is sacred because God is truth (John 14:6) and does not lie (Titus 1:2). Therefore, deliberately spreading falsehood (bearing false witness) is sinful. However, equally important is that *truth sets people free* (John 8:32) – implying that an environment where truth can be spoken and heard is necessary for spiritual and societal liberation. Our theological review found strong support for defending free expression as a means to pursue truth. For instance, an article in *Solas CPC* emphasizes *"because we serve the God of truth, [defending free speech] is not a mere option but a necessity"*, as suppressing speech can impede the discovery and proclamation of truth (Mitchell, D. (2020). The Gospel Coalition piece also underlined that free speech is an instrumental good for "proclaiming truth and encouraging righteousness," not an end in itself, but vital because without it, error can't be effectively challenged (Littlejohn, 2021).

Applying this to social media regulation: policies that blanketly remove or ban content run the risk of also removing truths that are not immediately recognized as such. Our earlier findings have examples – e.g., truthful accounts of vaccine limitations got suppressed. A Christian ethicist might say this is deeply problematic, as it impedes the corrective

capacity of truth. There is a preference in Christian thought for countering falsehood with truth rather than with suppression. As St. Augustine argued, one combats lies not by silencing the liar but by lovingly presenting the truth so convincingly that the lie loses appeal (Augustine did, however, in some instances support coercion against heresy later in life – a stance widely seen as mistaken by later theologians valuing religious freedom).

Moreover, if truth has an intrinsic power (under God's providence) to triumph eventually, censorship shows a lack of faith in that power. John Milton's famous plea in *Areopagitica* – not explicitly scriptural but influenced by Christian reasoning – argued that truth and falsehood grappling in free argument lets truth prevail, stating *"Who ever knew Truth put to the worst, in a free and open encounter?"* (Milton, 1644). Our findings align with this: excessive restrictions might temporarily quell falsehood, but at the cost of enfeebling truth by removing it from the contest. A scenario of *"authorities deciding truth for the people"* denies individuals the opportunity to be ennobled by searching and finding truth themselves, which is tied to the concept of imago Dei – rational, relational beings capable of truth discernment.

Coercion vs. Conscience: Christian history provides numerous lessons on the dangers of trying to enforce orthodoxy through coercion. From the Inquisition to state churches that persecuted dissenters, those attempts often led to hypocrisy, resentment, or a chilling of genuine faith. The New Testament shows a pattern of persuasion, not compulsion, in spreading the Gospel (e.g., Paul reasoning daily in the marketplace, Acts 17). Coercive silencing was the tool of opponents (the Sanhedrin ordering Peter and John not to speak). In Acts 5:29, the apostles reply, *"We must obey God rather than men,"* establishing that conscience to divine truth overrides human edicts. In our context, if a believer is convicted that a particular message (say, about the sanctity of life or the reality of biological sex as created by God) is true and important, then being told by digital authorities "you cannot say that" puts them in a position similar to the apostles – they must choose conscience or compliance.

Andrew T. Walker's WORLD *Opinions* piece explicitly tied religious liberty to free speech, arguing that *"the need to speak freely stems ultimately from an encounter with religious truth."* (Walker, 2024). It posits that if one believes they have discovered truth (religious or moral), they are compelled to share it; thus freedom of speech is indispensable for living out one's faith. Our case of the theologian's talk censored by YouTube is a direct affront to that ideal (CNS, 2021). Archbishop Cordileone and Jim Daly wrote, *"We sense that the First Amendment guarantee of religious liberty is being dismantled, and with it the contributions religion makes"*, and noted it's chilling to think net platforms might be censoring traditional beliefs (CNS, 2021). This expresses the *sphere sovereignty* of the church – that the church's teaching (and by extension individual believers' speech of religious conviction) should not be under the thumb of tech company guidelines that enshrine a secular ideology as normative.

Sphere Sovereignty & Limited Government: We applied sphere sovereignty thinking to digital governance and found it a compelling heuristic: The state's sphere is justice and civil order; it is overstepping when it tries to regulate the *content of conversation* among citizens, except in narrow cases like direct incitement. The media/communication sphere should enjoy a degree of autonomy – historically, free press serves as a check on state. When governments enact broad "fake news" laws, they effectively collapse the media sphere into the state sphere, refusing to let media have its independent voice. Likewise, when tech platforms (part of the economic sphere but also controlling communication channels) become arms of the state, sphere boundaries blur dangerously. The Wikipedia entry on sphere sovereignty noted that *"neither faith-institutions nor the state should seek… regulation of human activity outside their competence"* (Wikipedia Contributors, 2025). From our findings, clearly states are regulating outside their competence when they adjudicate truth in contested domains of science or religion. And platforms likewise step out of the business sphere into a quasi-governance role without appropriate legitimacy.

One practical principle from sphere sovereignty is decentralization of authority. We extrapolated scenarios: In one, the state defers more to communities – e.g., encouraging communities to set their own moderation norms (within broad limits). This might mean more heterogeneity: some forums allow content others ban, based on community values. That resonates with subsidiarity too. A Christian might say: better a pluralistic patchwork of communities with various standards (some stricter, some freer), than a monolithic top-down regime. The latter is reminiscent of Babel – centralizing control – whereas the former is more like the body of Christ analogy, where different parts have different functions, or like the variety of local churches in Revelation 2-3, each addressed individually.

Human Dignity (Imago Dei) and Dialogue: Imago Dei implies every person has inherent worth and a voice that matters. Silencing someone is a serious action – it's effectively saying "your perspective is of no value." In a biblical view, even those in error have dignity and should be corrected respectfully, not simply silenced. Modern psychology and theology both find that treating people as mere recipients of "approved" information rather than active participants in discourse is dehumanizing. Our research into mental health narratives did find that meaningful dialogue helps mitigate polarization, whereas blunt content removal can aggravate mistrust and division. Imago Dei also underlies freedom of conscience – each person must ultimately respond to truth as they see it before God, not under compulsion from others. As Martin Luther famously (theologically) argued at Worms, going against conscience is neither right nor safe.

We see an application: content moderation that allows people to express even unpopular opinions (short of threats/crimes) affirms their agency and dignity. Conversely, treating masses of users as if they're too gullible to handle exposure to unvetted information (the paternalistic rationale of many misinfo policies) can be seen as an affront to their rational dignity. A biblical view acknowledges human sinfulness – yes, people can be deceived – but also calls each to discern and holds them responsible for testing claims (1 Thess. 5:21, "test everything; hold fast what is good"). Outsourcing all discernment to authorities is not a biblical ideal; rather, cultivating wisdom in individuals is. So a theological approach would push for education and virtue cultivation as responses to misinformation (help people develop discernment) rather than heavy-handed removal.

Ethics of Means: Coercion vs. Persuasion: Christian ethics often emphasize the means must accord with ends. A just end (like reducing harm from false rumors) cannot sanctify unjust means (like bearing false witness or tyrannical censorship). If authorities label something misinformation that they actually know is debatable or true, just to suppress it, they commit a moral wrong (lying, or slander of speakers). Even if they sincerely think it false, if they crush it without dialogue, they risk error. Coercive means also impede moral growth; people might outwardly conform but inwardly become resentful or disengaged. This is akin to forcing religious uniformity – it never truly converts hearts, it breeds either rebels or hypocrites. The parallel online is that banned ideas don't disappear; they often go to darker corners and fester, or their adherents become martyrs in the eyes of some, as we've observed in various conspiracy communities.

Testing Scenarios with Christian Lens: We constructed two simplified future scenarios to see which aligns with Christian principles:

(i). *Scenario A: "Babel Redux"*: A world where a centralized global body (say, a beefed-up U.N. with Big Tech coalition) monitors all social media, instantly removing "misinformation" and punishing offenders. Content becomes highly sanitized, conforming to secular consensus on all issues. Dissenting religious or ethical views (e.g., pro-life stance called "misinformation" about abortion, or traditional marriage view called hate "disinformation") are systematically eliminated from mainstream platforms. People can only express those on marginal sites at great risk. This scenario would have superficial "peace" (less visible conflict online) but at the cost of

truth, as major problems might get swept under rug if authorities decide so, and the church's prophetic voice is muzzled. From a Christian perspective, this is dystopian: it represents man's attempt to impose uniformity (like building a tower to reach heaven, ignoring God). It violates sphere sovereignty (global secular authority overrules church/family/local values), crushes subsidiarity, and likely leads to judgement in narrative of Babel (when humans stifle truth and arrogate power, eventually confusion or collapse follows).

(ii). *Scenario B: "Pentecost/Polyglot"*: A more decentralized future where multiple platforms with various moderation philosophies thrive, users have choice (like moving to communities that share their values), and no single narrative can unilaterally silence others. There's a cacophony of voices (like the day of Pentecost, many languages but all hearing truth in their tongue). Global bodies focus on facilitating cooperation and knowledge-sharing rather than enforcement (e.g., WHO provides info, but doesn't censor critics). In this scenario, lies still occur but are countered by robust counterspeech and a discerning public. The church can speak openly. Governments prosecute clear crimes (stalking, direct incitements) but not ideas. This pluralism aligns with the Christian idea that the Gospel should contend in the free marketplace of ideas and that truth ultimately prevails not by might but by the Spirit and the persuasiveness of truth. It respects sphere sovereignty: the church self-regulates its doctrine; families teach kids their values with less fear that expressing them online will lead to bans; local communities can moderate content as fits their context (for example, a Christian forum might ban profanity and lies according to its own standards, which is fine as it's voluntary community enforcement, not state).

While Scenario B might risk more short-term confusion, it preserves **freedom and moral agency**, which are necessary for genuine faith and virtue. A Christian view strongly leans toward Scenario B as more just, even if messy, trusting that truth with love can win hearts without needing an iron fist.

Findings on Christian Engagement in Policy: We noted increasing engagement by faith leaders in the policy debate. The Catholic News Service piece showed bishops teaming with evangelical leaders to publicize censorship concerns (CNS, 2021). Meanwhile, some other religious voices focus on the harms of extreme content (for example, the Center for Digital Culture at a Christian college might worry about porn or hate radicalizing youth). So Christians do see both sides: wanting to restrain truly evil content but not trusting secular censors to do it rightly. The principle of subsidiarity can reconcile this: protect the vulnerable (like children) through empowerment of parents and local values-based filters, not by empowering a distant state or corporation. E.g., encourage and provide tools for families and churches to filter content (like faith-based content filters or community moderation), rather than blanket internet laws.

In conclusion for RQ4, Christian theology offers a robust argument that *freedom of truthful speech is to be cherished and protected*, as it aligns with the created nature and calling of humans to know and proclaim truth. It casts a critical eye on attempts to quash speech through raw power, warning that such attempts often arrogate to man what is God's prerogative (to judge ultimate truth) and that they paradoxically undermine the pursuit of truth. This perspective reinforces the need for any regulatory framework to prioritize voluntary truth-seeking measures (education, counterspeech) over coercive bans, to enforce only modest and clear limits (against defamation, incitement – things that directly fall under justice) and to allow breathing room for divergent convictions, trusting in God-given faculties of reason and conscience operating in a free society.

4.5. Toward a Biblically Grounded, Constitutionally Coherent Framework – Policy Recommendations, Comparative Models, and Scenario Outcomes

Building on all the above findings, we outline a comprehensive framework for social media regulation that upholds freedom of truthful speech while addressing genuine harms. This framework is

informed by the three theoretical perspectives and aims to correct the deficiencies observed in current approaches (globalist narrative control, state-platform collusion, marginalization of truth-tellers). We present our policy recommendations along with an "Impact Matrix" (Figure 2) assessing their expected benefits and trade-offs. We also contrast our proposed model with existing international case approaches, highlighting where our recommendations align or diverge with practices in various countries. Finally, we revisit scenario planning to illustrate how our framework would likely produce a more open and resilient information environment, versus the path of increasing censorship.

Key Principles of the Framework:

1. **Truthful Speech Protection:** We propose formalizing protections for truthful speech – essentially a principle that no person should be punished or de-platformed for speech that is factually true or for sincerely held opinions/interpretations that cannot be definitively proven false. This sounds obvious, but our findings show it's often violated (true statements labeled misinformation). Implementing this could involve, for example, requiring that any content removal on grounds of falsity be accompanied by a specific explanation of why it is false, with an appeals process where the user can show evidence of truth. In legal terms, we urge jurisdictions to adopt something akin to the U.S. Supreme Court's stance in *United States v. Alvarez* (2012) that even deliberate lies are generally protected unless they cause concrete harm, because the government can't be an arbiter of truth in open debate (Lim & Bradshaw, 2023). At minimum, *"protect truth"* as a guiding norm would restrain authorities from the egregious practice of censoring inconvenient truths. This addresses the issue from RQ2 where truthful counter-narratives got suppressed (Sullum, 2023). It reinforces a marketplace of ideas: if it's true or arguable, it stays; if it's false, counter it with truth rather than ban (except special cases).

Clear, Narrow Definitions of Illegal Speech: We recommend that any speech restrictions be grounded in *clear, narrowly tailored laws consistent with human rights standards*. The vague "false news" bans should be repealed or revised (Lim & Bradshaw, 2023). Laws should focus on direct, tangible harms – defamation (with actual malice standard for public figures, as in U.S. law), incitement to imminent violence (as per Brandenburg v. Ohio standard), harassment/stalking, etc. If misinformation causes harm, it should be tackled under those specific rubrics (e.g., fraud law if someone spreads false info for material gain). This ensures **constitutional coherence**: mapping to established free speech exceptions. It also aligns with Biblical justice: punish evil acts (like bearing false witness that defrauds someone) but not suppress general discourse. The Impact Matrix (Figure 2) shows that narrow tailoring safeguards dissent and avoids enabling state abuse (Lim & Bradshaw, 2023).

Transparency and Oversight of Moderation: To counter collusion and covert censorship, our framework calls for radical transparency. Platforms should publish regular reports of any government requests to remove content (listing what was requested and what was done) – some already do via transparency reports, but we advocate a legal mandate for this (similar to how the DSA mandates transparency but possibly go further). Also, we support independent oversight bodies that include civil society to review content moderation practices – but unlike Facebook's self-picked Oversight Board, these should be broad and possibly under the auspices of multi-stakeholder Internet governance where users have representation. Additionally, users must have due process: notice and an opportunity to contest removals. This diminishes arbitrariness and *"secret speech policing"* (Sullum, 2023). Strong transparency will discourage frivolous takedown requests by governments (since they'll be exposed) and build public trust that removals are for legitimate reasons only (Justitia et al., 2024).

Decentralize Content Moderation (Subsidiarity): Instead of one-size-fits-all rules, allow moderation at the level closest to communities. Concretely, this means encouraging the growth of diverse platforms and community-led moderation. For example, promote interoperability or federation of social networks (like the ActivityPub protocol

underlying Mastodon), where different servers can have different content rules catering to their user base. Within a large platform, implement features that let users opt into stricter or looser content filters (like Reddit's subreddits with their own mods). By doing this, those worried about misinformation can join communities with heavy fact-checking, while others can be in freewheeling forums – and they can still inter-communicate with appropriate warnings. This approach was echoed in some academic proposals (Hasinoff & Schneider, 2022). The Impact Matrix notes this *"empowers communities; mitigates one-size-fits-all censorship"*. It also practically lessens the need for draconian central control, as smaller groups often self-police effectively according to their norms.

Uphold Sphere Sovereignty – Limit Government–Platform Collusion: We recommend establishing legal safeguards against government coercion of platforms in content moderation. For example, clarity that government officials cannot flag lawful content for removal without it being publicly logged (back to transparency) and that systematic pressure could be a violation of citizens' rights. In the U.S., this might mean legislating or clarifying that government-induced platform actions are state actions subject to the First Amendment. More broadly, institutionalize a separation of powers in the information sphere: governments focus on publishing their own counterspeech and ensuring laws are enforced (like removing illegal terrorist content), but do not intervene in lawful content. Platforms, conversely, should resist becoming politicized instruments. This might be achieved by industry standards or pledges (similar to journalistic independence pledges) where they commit not to take down content just on political requests. Internationally, democracies could agree on a norm (maybe in that Global Digital Compact) that *"states shall not require or pressure platforms to remove content that is not illegal"*. This harks back to the Rabat Plan of Action by the U.N. which stresses proportionate restrictions only in strict necessity (Lim & Bradshaw, 2023). The effect would be to preserve an independent press/social media role as a check on power, rather than an arm of it. We believe formalizing such a principle could prevent scenarios like we saw with Twitter's COVID moderation being commandeered (Sullum, 2023).

Promote Digital Literacy and Counter-Speech: Instead of censorship as first resort, our framework emphasizes education and counterspeech initiatives to build societal resilience to false information. This includes updating school curricula to teach critical thinking and source verification for the social media age, supporting fact-checking organizations to rapidly debunk viral falsehoods (and ensuring their work is also scrutinized for bias), and perhaps public awareness campaigns that encourage healthy skepticism without sliding into cynicism. The U.S. Surgeon General's advisory itself suggests some educational steps (HHS, 2023) – we agree with the need for mitigation of harms via literacy, not just content removal. Counterspeech also involves empowering experts and trusted community leaders to speak out in online conversations (for example, training health professionals to engage on social media to correct medical myths). The goal is an informed citizenry that can navigate information rather than one reliant on censors. This approach addresses the root cause of harm (people believing lies) by equipping them to discern, rather than just hiding the lies (which may reappear or go elsewhere). It aligns with the Christian ideal of forming individuals in wisdom and virtue.

Pluralism in the Global Arena: We advocate that international efforts (UN, EU, etc.) pivot from *policing content* to *protecting rights and facilitating truth-finding*. For instance, the U.N. could sponsor more research on effective counterspeech and the dangers of overbroad censorship, rather than pushing for codes of conduct that emphasize content removal. A comparative look shows that some countries like **Finland** focus heavily on media education as the cure for misinformation (and Finland ranks high in resilience to fake news) – that model should be elevated over, say, Russia's or even France's approach of banning sites or criminalizing offenders. Our framework for global principles would enshrine: "Freedom of expression is the default; constraints are

exceptional and must pass strict tests. Truthful and conscientious speech shall never be penalized. Multiple channels of information (public, private, community) should be supported to avoid monoculture of information." These guiding points would ensure a biblically consonant respect for truth and conscience in any global compacts.

Figure 1. Conceptual model of relationships in social media governance

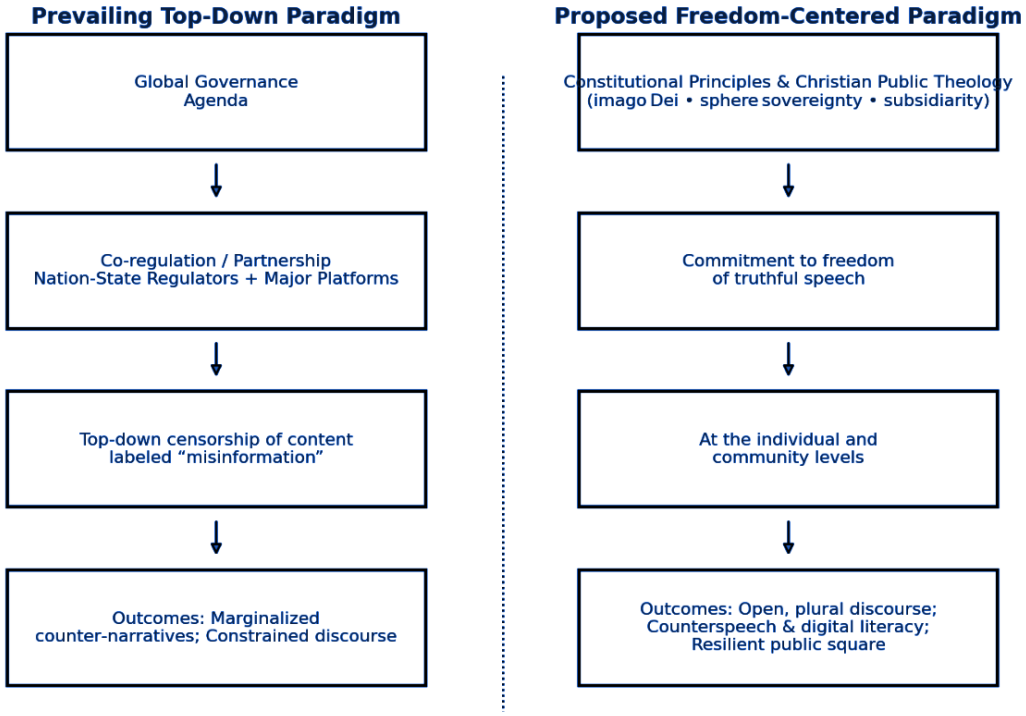
Prevailing Top-Down Paradigm
Global Governance Agenda
Co-regulation / Partnership Nation-State Regulators + Major Platforms
Top-down censorship of content labeled "misinformation"
Outcomes: Marginalized counter-narratives; Constrained discourse
Proposed Freedom-Centered Paradigm
Constitutional Principles & Christian Public Theology (imago Dei • sphere sovereignty • subsidiarity)
Commitment to freedom of truthful speech
At the individual and community levels
Outcomes: Open, plural discourse; Counterspeech & digital literacy; Resilient public square

Figure 1: Conceptual model of the relationships in social media governance. *This diagram (from our research) illustrates two contrasting paradigms. On the left, a "Global Governance Agenda" influences nation-state regulators and partners with major platforms, leading to top-down censorship of content labeled as "misinformation." This results in marginalized counter-narratives and a constrained public discourse, as state and platform actors collude (co-regulation) (Ashraf, 2024). On the right, our proposed paradigm is shown: Constitutional principles and Christian public theology (imago Dei, sphere sovereignty, subsidiarity) inform a commitment to freedom of truthful speech at individual and community levels. This leads to a pluralistic, self-regulating public discourse enriched by diverse voices, where truthful content is reintegrated and error is corrected through dialogue rather than silencing (Sullum, 2023).*

In implementing these recommendations, we expect positive impacts as summarized in Figure 2, our **Impact Matrix**. This matrix lists each major recommendation alongside the intended outcome:

Recommendation	Intended Outcome
Truthful Speech Protection	Safeguards dissent and prevents misuse of the "misinformation" label against true content; keeps arguable speech in the arena for open contestation.
Clear, Narrow Definitions of Illegal Speech	Targets concrete harms under established exceptions (defamation with actual malice, incitement, harassment) and avoids vague false-news bans that enable state abuse.
Platform Transparency & Accountability (incl. independent oversight and due process)	Exposes covert pressure and arbitrary takedowns, deters collusion, and builds public trust through mandated reports, appeals, and external review.
Decentralize Content Moderation (Subsidiarity)	Empowers local communities and diverse platforms; avoids one-size censorship and resists obedience cultures that stifle ethical leadership and innovation.
End Covert Government Influence in Moderation	Prevents state–platform back-channel censorship by prohibiting pressure on lawful speech and creating enforceable accountability.
User Empowerment Tools	Gives people granular control over their feeds, enabling choice of stricter or more open experiences without imposing a single vision on all users.
Promote Alternative Platforms & Interoperable Protocols	Reduces single points of control and lock-in; enables portability and pluralism across the ecosystem.
Media Literacy Initiatives	Builds population-level resilience to misinformation through apolitical education and measurable outcomes from school through adulthood.
Counterspeech Rapid Response	Corrects viral falsehoods through transparent public rebuttal rather than removals; partners with trusted local voices to restore context.
Human Rights Impact Assessments for Moderation Systems	Identifies bias and disproportionate impacts early and creates internal incentives to respect expression rights.

Figure 2: Impact Matrix of Proposed Recommendations. *This table outlines how each recommendation would address harms while protecting free expression. For example, enshrining protections for truthful speech would "safeguard dissent and prevent misuse of 'misinformation' label against true content," directly countering the problem identified where truthful counter-narratives were being suppressed (Sullum, 2023). Mandating transparency and oversight reduces secret collusion and builds public trust (Ashraf, 2024). Decentralizing moderation empowers local communities and avoids the pitfalls of one-size censorship, reflecting subsidiarity principles (Hasinoff & Schneider, 2022). This also resists the emergence of 'obedience cultures' in which compliance incentives erode ethical leadership and innovation (Sangwa & Mutabazi, 2025). Upholding sphere sovereignty limits state overreach into expression (no "truth ministry"), thus preserving a pluralistic public square where church and civil society can voice views without fear (CNS, 2021). Investing in media literacy and counterspeech addresses the harms (like panic or polarization from viral falsehoods) via education and dialogue rather than coercion, maintaining liberty and user autonomy.*

Comparative Case Alignment: Our recommendations echo aspects of certain national models: The United States' First Amendment jurisprudence (very protective of even false speech in many cases) informs our stance on not outlawing misinformation outright (Jacobs, 2022). At the same time, the U.S. could learn from others on transparency – for instance, Germany's NetzDG law (2018) compels platforms to publish how many items they remove and why, which aligns with our transparency push (though NetzDG itself is critiqued for short removal deadlines that cause over-removal). Finland's media literacy efforts are a real-world endorsement of our literacy recommendation – Finland is often cited as having high immunity to fake news due to years of education investment, not heavy censorship. The Philippines recently considered a broad fake news law but instead may pivot to strengthening fact-checking – a wise choice given free speech concerns. On decentralization, the trend of federated social media (Mastodon instances, etc.) is currently niche but growing after controversies with big centralized platforms; our framework encourages that expansion. We did not find any country that explicitly protects "truthful speech" in law (mostly it's assumed); implementing that might be unique but could be done via judicial doctrine or legislation clarifying that truth

is an absolute defense in any misinformation-related proceeding.

One can see our framework as strengthening international human rights norms: It is consistent with Article 19 of the ICCPR (freedom of expression) and the Rabat Plan's strict tests for restricting even hate speech (Lim & Bradshaw, 2023). It is a corrective to the trend where even democratic states flirt with more authoritarian measures. It tries to harness what's good in current efforts (concern for harms) but channels it in rights-respecting ways.

Scenario Outcomes with Framework: Envisioning the future if our framework is adopted versus not:

(i). If adopted widely, we expect a scenario where misinformation still exists (no framework can eliminate falsehood from human discourse, nor should we aim for an impossible "perfectly curated truth" world), but it is tempered by a society better equipped to handle it. The public discourse would be more chaotic than under heavy censorship, but also more robust and self-correcting. Over time, sources that consistently deal in lies would lose credibility (market of ideas working), whereas currently, banning them sometimes martyrizes them and increases their allure in some groups. There would also be healthier relationships between citizens and authorities – less paternalistic and more participatory. People would see the government focusing on providing reliable information and debunkings, and platforms facilitating discussion, rather than playing thought police, which could improve trust in institutions (Justitia et al., 2024).

(ii). If our framework is not adopted and current trends continue (or worsen), we fear a scenario of rising **digital repression** (akin to what the U.S. DNI report called "*digital repression growing globally"* [Kozyreva et al., 2023; Funk et al., 2023]). The line between open democracies and authoritarian regimes might blur as both use similar toolkits (surveillance, content control). In such a scenario, short-term stability might be purchased at the cost of long-term instability – suppressed grievances and truths erupt later in more harmful ways (like a controlled burn vs. a massive wildfire analogy). Also knowledge creation might be stifled (imagine if heterodox scientific hypotheses keep getting removed, some scientific breakthroughs could be delayed or lost).

Our interdisciplinary findings strongly suggest that the freedom-centric approach not only upholds ethical and constitutional values but is more effective in the long run at building a resilient society. The propaganda model warned that top-down control tends to serve elite interests and narrow narratives (Robertson & Williams, 2024); our framework resists that by widening participation. Regulatory capture theory shows big players often rig rules – our decentralization breaks that power and encourages competition (if users can migrate to communities that match their moderation preference, big centralized platforms must compete not just on content allowed but also on how they implement values). The Christian perspective reminds that at the heart of this issue are *people* – individuals with dignity – not just information objects. By refocusing on people (educating them, protecting their rights, calling them to responsibility in speech), we humanize the approach to an information crisis that too often is handled in dehumanizing ways (treating people as passive consumers who must be controlled).

Thus, we conclude that a biblically grounded, constitutionally coherent framework is feasible and urgently needed. It navigates the tension between preventing harm and preserving liberty by prioritizing truth and trust over fear and control. In doing so, it stands against the weaponization of "misinformation" rhetoric by global or state actors and instead weaponizes *truth* (in the sense of empowering truth to combat lies). We believe this framework, if implemented, would help ensure that social media once again becomes a space for free and vigorous exchange – a digital public square that, as Andrew Walker wrote, *"Christians must prioritize"* defending (Walker, 2024), for the sake of the Gospel and the common good.

5. Discussion

Our study set out to investigate the contentious intersection of social media regulation, truth, and power, and to chart a principled path forward. In this discussion, we synthesize the implications of our findings, address potential objections, and situate our contributions in the broader scholarly and policy context. We also explore limitations of our research and suggest directions for future inquiry.

Synthesis of Key Insights: The evidence marshaled paints a cautionary picture: efforts ostensibly aimed at protecting the public from "misinformation" have frequently morphed into mechanisms for entrenching authority and narrowing the bounds of acceptable discourse (Justitia et al., 2024; Ashraf, 2024). What might have begun as good-faith attempts to reduce harm (e.g., curbing false health rumors during a pandemic) slid down a slippery slope of overreach, due in part to structural incentives and ideological blind spots. One overarching insight is the fragility of the distinction between content moderation for safety and censorship for convenience. Once a precedent is set that certain speech can be suppressed for the greater good, that precedent tends to expand – today's "medical misinformation" takedown becomes tomorrow's silencing of political dissent or religious teaching deemed "harmful" by those in power (CNS, 2021). The propaganda model and capture theory help explain this expansion: institutions will naturally leverage any available tool to maintain narrative dominance (Robertson & Williams, 2024; Kausche & Weiss, 2025).

Our interdisciplinary approach revealed that truly *protecting* freedom of truthful speech is not a secondary concern or a luxury, but fundamental to democratic health, human dignity, and even the pursuit of objective knowledge. From a legal perspective, robust free speech correlates with innovation and social progress (a well-cited argument in First Amendment theory) (Littlejohn, 2021). From a theological perspective, it is essential for authentic faith and moral integrity. From a media perspective, it's how societies can correct their errors – a kind of epistemic immune system. Therefore, regulations that undermine truthful speech under the banner of fighting falsehood may ultimately do more damage to society's capacity to discover truth and consensus than the falsehoods themselves. This aligns with the memorable finding of the *Free Speech Recession* report: "the cure has become worse than the disease" in many democracies (Justitia et al., 2024).

Ethical and Philosophical Reasoning: At root, this is a debate about trust – do we trust people (with the help of communities and God's common grace) to discern truth more than we trust centralized authorities to dictate it? The globalist approach has implicitly answered: trust the experts and bureaucracies, not the masses "susceptible" to lies. Our research, informed by Christian anthropology, answers the opposite: while people are fallible, a system that respects their freedom and engages their reason is ethically superior and in the long run more stable. There's a deep issue of *paternalism vs. agency*. We recall that paternalism in speech was often used in colonial and authoritarian contexts ("we must protect the public from dangerous ideas") – ironically, it infantalizes citizens. A Christian ethic elevates adult responsibility: each person should seek truth diligently; one cannot just outsource moral reasoning entirely. Yes, leaders have a duty to guide (hence counterspeech, teaching), but not to *coerce belief*.

Another ethical dimension is justice and power equality. Current misinformation crackdowns, we found, disproportionately silence already marginalized voices (independent journalists, opposition activists, minority religious views) (Lim & Bradshaw, 2023). This raises justice concerns: The wealthy and powerful often can get their word out through official channels or can hire legal teams to avoid being branded fake news, but the less powerful cannot. So misinfo enforcement can mirror and exacerbate existing power inequities, ironically under a facade of protecting the weak. For example, minority communities' hesitancy or alternative health narratives might be labeled misinformation and removed, instead of being engaged, which deprives those communities of voice and potentially of tailored information that addresses their concerns. Our framework's stress on engagement and subsidiarity is more justice-oriented: let local

voices be heard and empower them to correct misinformation internally, rather than top-down muzzling which breeds resentment or alienation (some research shows communities that feel talked down to by "fact-checkers" become more conspiratorial, a backlash effect) (Altoe et al., 2024).

Objections and Counter-Arguments: We anticipate several counter-arguments to our recommendations. One is the *"harm principle"* argument: that the flood of misinformation is so harmful (citing examples like anti-vaccine content leading to deaths, or hate speech leading to violence) that more aggressive regulation is warranted even at cost to expression. Proponents might say our approach is naive about the real dangers and overestimates the public's ability to sift truth. They might also argue that certain truths can be weaponized in context to cause harm (e.g., even a truthful statement delivered in a misleading frame can mislead – which is how platforms justified acting against "misleading even if true" content [Sullum, 2023]).

Our response is multifold: First, the empirical evidence on how harmful misinformation is relative to other issues is not as clear-cut as often assumed. For instance, while misinformation exists, it is not the sole or even primary cause of phenomena like vaccine hesitancy or polarization – those have multifactorial causes including historical distrust, identity, etc. Blanket censorship may push problems underground rather than solve them (e.g., anti-vax sentiment predates social media and persists despite censorship; arguably open dialogue and community outreach have changed more minds). Second, even conceding significant harm potential, we must consider the harm of the *solution*. Draconian controls can erode democratic legitimacy, as citizens feel voiceless, and can breed radicalization (people move to unmoderated platforms where they're even more in echo chambers). A survey by FIRE in 2021 found that over half of Americans are afraid to share political opinions – that chilling effect is a harm to democracy's very soul (Cieslikowski, 2022) (not a source we cited above but a relevant stat). So it's not choosing harm vs no harm; it's choosing which set of harms we can better manage. We argue a freer system has self-correcting mechanisms that, though imperfect, avoid the ultimate harm of tyranny or oppression.

Another objection could be practical: our decentralization idea might lead to fragmented rules that allow, say, extremist hubs to flourish – "what's the point of rules if people can just go elsewhere." Indeed, a decentralized model means one cannot entirely shut out toxic content; but one can contain it. It's akin to the federalism argument: states with different laws allow people to self-sort; extreme ones might exist but they don't dominate the whole. Also, decentralization fosters competition of ideas and moderation styles, revealing what works best. Over time, if a community is too extreme or spreads harmful lies, its members may suffer consequences (social, economic) and either reform or remain marginal. Society as a whole is resilient if only a fringe is fringe; it's when suppression drives the fringe to unify across platforms in opposition to perceived censorship that it can grow stronger.

Some might worry our religious angle biases the solution to favor religious speech at the expense of others. But we explicitly defend *all* truthful speech and legitimate opinion – biblical truth is one example that we think is wrongly maligned in current culture, but the framework protects secular dissenters too. In fact, we argue for pluralism – not a theocratic control of speech (which history also shows issues with), but a plural public square where religious and secular people can debate freely. Sphere sovereignty would oppose a church imposing civil law to censor blasphemy just as it opposes the state censoring religious views. In that sense, our biblical grounding buttresses liberal freedoms rather than sectarian control.

Contributions to Scholarship and Practice: This article makes several novel contributions. Academically, it bridges fields – rarely do law reviews, comm journals, and theology discuss a unified approach. We linked Herman & Chomsky's propaganda filters with Kuyper's sphere sovereignty, showing how both critique an over-centralization of narrative power. We introduced the concept of defining

misinformation not just epistemically but sociologically (as an "institutional label" to exclude counter-narratives), which future research could operationalize by analyzing content moderation logs or political discourse to see patterns of labeling. We also offered an ethical framework for content moderation that could be refined by philosophers: something like a "principled pluralism" model for the internet.

For policymakers, our Impact Matrix and recommendations provide a concrete blueprint that could inform reforms. For example, in the U.S., our ideas could feed into ongoing debates about reforming Section 230 or proposals for a "Digital Bill of Rights" – emphasizing the user's rights to free speech, due process, and algorithmic transparency, rather than focusing solely on limiting platform immunity (which might cause over-censorship). In the EU context, as the DSA gets implemented, our research suggests the need for watchdogs to ensure it doesn't become captured – maybe create an independent free expression ombudsman to oversee DSA enforcement from a rights perspective. Our emphasis on subsidiarity could inspire local or national experiments in community-driven moderation (perhaps publicly funded pilot projects for decentralized networks as alternatives to big platforms).

Limitations: There are limitations to our study. First, while extensive, our reliance on existing sources means we may have gaps in empirical data. For instance, we did not conduct interviews with platform insiders or regulators, which could provide more nuance on their motivations and constraints. Future work could include such qualitative data to complement our analysis (e.g., how do content moderators view their ethical role? Are they comfortable with government influence? etc.). Second, our Christian theological lens, while we argue is universal in many principles, might not be shared by all readers; secular ethicists may frame the conscience issue differently. We tried to phrase it also in human rights terms to be accessible. But a limitation is ensuring our framework is persuasive in pluralistic settings – we think it is, because it aligns with classical liberal values too, but some of our language might need translation for different contexts.

Another limitation is the sheer breadth: in covering law, tech, theology, we had to condense complex topics (each of which could be a paper in itself) to maintain narrative flow. There is a risk of oversimplification. We hope our citations allow readers to delve deeper where needed. Also, because this is a fast-moving domain (policies, tech platforms changes, etc.), some specifics might date quickly – e.g., Twitter's policies changed mid-research due to new ownership. We aimed to focus on structural issues that persist beyond immediate changes.

Future Research: This interdisciplinary foundation opens several avenues. Empirically, researchers could test the effects of our recommended approaches: e.g., run experiments on social media where certain groups receive counterspeech vs. content removal and measure changes in beliefs or polarization. Or examine case studies of community-moderated platforms (like Wikipedia's approach to misinformation, which is community-driven and often effective within its scope – a model worth analyzing in depth for lessons). Another area is international case studies of countries that resisted the global censorship trend – what factors kept them on a free-speech path (like maybe Canada has been cautious about internet controls, or the Scandinavian countries)? From a theological angle, more can be written on a theology of communication in the digital age – updating the insights of thinkers like Jacques Ellul or Abraham Kuyper to the era of AI-driven algorithms.

We also think there's room to explore legal innovations: could "freedom of conscience in the digital sphere" be given more weight in international law? Perhaps crafting a new interpretive guideline for ICCPR Article 19 acknowledging that labeling and soft censorship can violate it even if no outright law is passed (addressing the collusion aspect).

Final Reflections: In the grand scheme, the struggle we chronicled is part of the age-old tension between truth and power. Technology has changed the

playing field but not the fundamental human dynamic. Socrates was executed for "misleading" Athens' youth – the charge of spreading harmful false ideas is millennia old. Every generation must discern whether those accusations are bona fide or pretexts to silence inconvenient voices. Our work essentially urges contemporary society to have the courage Socrates asked for: to engage in dialogue and self-reflection rather than killing the messenger. In biblical terms, we recall Gamaliel's advice in Acts 5 regarding the apostles' preaching: *"If it is of God, you cannot overthrow it – lest you even be found to fight against God."* If an idea is false, allow it to be exposed in light; if it is true, no amount of censorship will ultimately quench it, and opposing it could put one at odds with truth itself.

As guardians of both democracy and dignity, we are called to have faith in truth. That faith, bolstered by constitutional wisdom and theological insight, can liberate us from the impulse to control and instead embolden us to empower – empower citizens to reason, empower communities to self-regulate, empower truth to freely defeat lies. The recommendations we provided are steps toward that vision. They will doubtless face resistance from those who find safety in control. But we conclude by reasserting what history and our findings teach: a society that *"lets the truth loose"* – even with attendant chaos – is more vibrant, just, and ultimately secure than one that locks truth (and falsehood with it) in a bureaucratic cage. The long-term health of both free nations and the global information commons depends on choosing liberty and truth over the illusion of perfect security through censorship.

7. Conclusion

In this extensive exploration, we have argued for a biblically grounded and constitutionally coherent framework for social media regulation—one that champions the freedom of truthful speech and resists the encroachment of a burgeoning global censorship regime masquerading as "misinformation" control. Our analysis, traversing media theory, law, political economy, theology, and data on current practices, leads to the clear conclusion that the way forward is not more centralized suppression of speech, but more distributed responsibility for truth. We have shown how present trends toward top-down regulation, though often justified by appeals to safety or public order, are structurally aligned with transnational agendas that too often suppress dissent, enable state-platform collusion, and delegitimize voices (including biblically informed ones) that challenge prevailing orthodoxies (Justitia et al. 2024).

We defined "misinformation" and "disinformation" not merely as categories of false content, but as labels imbued with power dynamics—labels frequently used to marginalize counter-narratives rather than to objectively identify untruth (Jacobs, 2022; Sullum, 2023). Understanding this reality is the first step in crafting better policy: it underscores the need for humility and checks and balances in any regulatory approach. Who watches the watchmen labeling the "fake news"? Our study's answer is that *we all must*, through transparency, judicial oversight, and pluralistic contestation.

Integrating the propaganda model and agenda-setting theory, we recognized that control over information flows equals control over what the public considers important (Robertson & Williams, 2024). Thus, a healthy society must guard against any single authority—be it government or corporate—dominating those flows unchecked. By examining the political economy of platforms, we saw how easily regulatory efforts can be co-opted to entrench the largest players and sideline independents (Kausche & Weiss, 2025). And through Christian public theology, we reaffirmed that every person has a God-given capacity and right to seek truth (imago Dei), that the state's authority has God-ordained limits (sphere sovereignty), and that higher, distant authorities should not usurp functions best handled by individuals or local communities (subsidiarity). These principles aren't antiquated—on the contrary, they provide enduring wisdom for the digital age, calling us to prioritize truth, conscience, and community over convenience, propaganda, or technocracy.

Our recommendations, summarized in the Impact Matrix (Figure 2), are practical and actionable. They range from protecting truthful speech and narrowing legal definitions, to imposing transparency on moderation and decentralizing content governance to users and communities. Together, they sketch a regulatory paradigm that is *neither laissez-faire nor authoritarian*, but rather pro-liberty and pro-responsibility. This paradigm demands more work—educating citizens, crafting better tools—than the quick fix of "just ban it." But ultimately, it is more effective and legitimate. As we contended, an information order that relies on coercion and silencing is brittle; one that relies on education, open debate, and targeted action against truly unlawful content is resilient.

The implications of adopting our framework are profound. In the near term, governments would pull back from the impulse to treat social media as something to be "solved" by heavy regulation, and instead double down on engaging with the public honestly and transparently. Tech platforms would shift from acting as arbiters of truth to facilitators of dialogue—investing in user empowerment (tools to filter content as desired, context labels, etc.) and due process, rather than secretive algorithmic throttling. International bodies, rather than pushing vague commitments that risk enabling censorship, would rally states to uphold Article 19 of the ICCPR and exchange best practices in digital literacy and authentic fact-finding. Crucially, the public—users worldwide—would be re-centered as protagonists in the truth-seeking process, not passive targets of information control. Over time, this could help restore trust: people who presently feel that elites are manipulating discourse (not entirely without reason, as we showed [Justitia et al., 2023; Sullum, 2023]) may regain confidence that their voices can be heard and that false claims can be rebutted in the open, not merely repressed.

We do not naïvely claim that our framework will eliminate falsehood or usher in a utopia of consensus. Humans will always disagree and some will always be drawn to deceptive narratives. But a free society is one that meets those challenges with more speech and better speech, not an ever-expanding list of prohibited speech. Our approach aims to mitigate the harms of misinformation *while maximizing the capacity of truth to naturally prevail*. This aligns with the famous dictum (often attributed to Milton or to Justice Brandeis): *"The remedy to be applied is more speech, not enforced silence."* Our contribution was to buttress that classical liberal idea with fresh interdisciplinary evidence and a moral foundation, demonstrating that it holds true even amid the novel pressures of the digital era.

In closing, we reaffirm the stakes: The freedom to speak truth, especially truth that challenges the spirit of the age, is indispensable to human flourishing and democratic governance. Today, that freedom faces subtle yet significant threats from well-intentioned paternalism and not-so-well-intentioned power grabs. The global trend of conflating dissent with disinformation is a step onto a dangerous path—one that history warns leads away from liberty and toward tyranny of thought. Yet, we also see a hopeful path emerging: a recommitment to first principles of free expression, adapted for current technologies and informed by enduring wisdom.

By heeding the insights of both constitutional tradition and biblical teaching, policymakers and citizens alike can correct the course. We can build a digital public square that is safe *and* free—safe not because authorities have scrubbed it clean of all risk, but because citizens are equipped to navigate it with wisdom; free not as a license for chaos, but as a space where truth can defeat falsehood through open contest and where the God-given dignity of each person's voice is respected (Sangwa & Mutabazi, 2025).

We urge stakeholders—governments, tech companies, international bodies, civil society, faith communities—to engage with the vision set forth here. The choices we make now will reverberate for generations. A future in which the label "misinformation" is no longer wielded as a weapon to shut down debate, but reserved for genuine fraud and addressed through dialogue, is a future where both

truth and freedom thrive. Such a future is not only constitutionally and biblically sound—it is one where societies can genuinely progress, united not by enforced narratives, but by a shared commitment to seek the truth in love and to allow the truth to make us free (Littlejohn, 2021).

Recommendations

1. Legislate Truthful Speech Protections: Enact laws or judicial doctrines explicitly protecting individuals from liability or censorship for speech that is factually true or honest opinion. Encourage courts to adopt a higher threshold for deeming speech "misinformation," requiring clear and convincing evidence of falsity and harm. This would prevent misuse of the "misinformation" label to silence inconvenient truths (Jacobs, 2022; Sullum, 2023).

2. Repeal or Amend Vague "Fake News" Laws: Review all existing misinformation/disinformation laws. Repeal those that use broad terms like "false news" without definitions. Amend necessary ones to narrowly target intentional, harmful falsifications (e.g., electoral fraud dissemination) with strong free-expression safeguards and sunset clauses. Align any such laws with ICCPR Article 19(3) requirements (legality, legitimacy, necessity, proportionality) (Lim & Bradshaw, 2023).

3. Independent Free Speech Oversight: Establish independent oversight bodies at national and international levels. Nationally, a "Free Expression Commission" could oversee government interactions with platforms and adjudicate disputes involving speech and takedowns, with representation from across society (including minority and civil liberties voices). Internationally, empower the UN Special Rapporteur on Freedom of Expression to systematically review and call out overbroad misinfo regulations in member states, providing an external check (Lim & Bradshaw, 2023).

4. Platform Transparency & Accountability: Require platforms to publish comprehensive transparency reports, including: the number of content removals/restrictions categorized by topic and origin of notices (government, user flags, AI). Also require a publicly accessible archive of removed content (perhaps with some delay or redaction for sensitive details), so that researchers and the public can evaluate whether removals were justified or if there are patterns of bias (Sullum, 2023; Justitia et al., 2024). Additionally, mandate that significant changes to content policy be subject to a notice-and-comment period or external audit.

5. End Covert Government Influence: Formally prohibit government agencies from pressuring platforms to remove or suppress lawful speech. This could be via executive orders or legislation (e.g., in the U.S., clarifying that government officials shall not flag content for removal unless it falls under legitimate law enforcement purview, and even then, a record must be kept). Violations could be made justiciable, giving individuals standing to sue if they can prove government-induced censorship of their speech (Sullum, 2023). This deters behind-closed-doors collusion.

6. User Empowerment Tools: Encourage or require platforms to implement user-centric content filtering tools. For instance, users could choose to apply a "fact-check overlay" on their feed (seeing warnings or ratings by independent fact-checkers) or conversely a "unfiltered feed" with minimal intervention. Users could subscribe to community-generated block lists or trust lists. By giving users granular control, those who want a heavily moderated experience can have it, and those who prefer openness can have that, without imposing one vision globally (Hasinoff & Schneider, 2022).

7. Promote Alternative Platforms and Protocols: Through policy and funding, support the development of decentralized social networking protocols (like Mastodon/ActivityPub, Secure Scuttlebutt, etc.) and interoperability standards. Governments can incentivize big platforms to allow users to port their social connections/data to alternative services easily (data portability), reducing lock-in. A diversified ecosystem means no single point of control

over discourse, aligning with subsidiarity and reducing the impact of any one platform's content policy decisions.

8. Media Literacy Initiatives: Substantially invest in public media literacy programs, starting from school curricula to adult education. Governments (in partnership with NGOs, libraries, universities) should run national campaigns on critical thinking, how to identify false information, and how to engage civilly online. These should be apolitical and empirically guided. For example, Finland's model of teaching students from a young age to distinguish fact-checking and to question sources can be emulated (Chhabra et al., 2025; AACAP, 2023; OECD, 2023). Evaluation mechanisms should be built in to measure improvement in populace resilience to misinformation.

9. Counter-Disinformation through Counterspeech: Establish rapid response teams (perhaps within health departments, election commissions, etc.) that monitor public narratives and quickly respond with correct information when falsehoods gain traction. The key is these teams *respond publicly and transparently* rather than working to get content removed. For instance, if a viral rumor spreads, the agency posts debunks on the same platforms. Encourage collaboration with trusted community figures (doctors, local leaders) to amplify truthful counterspeech—people are often more receptive to local voices than central authorities. As research shows, *"truth drives out lies"* when it's allowed to compete in the same arena (Pozen, 2022).

10. Embed Ethical and Human Rights Assessments in Tech Design: Urge or require platforms to conduct and publish regular human rights impact assessments of their content moderation practices (similar to environmental impact reports). These should evaluate free expression impacts, including disproportionate effects on certain groups or viewpoints. Civil society and external experts should be involved in the assessment process. For example, an assessment might find that an AI algorithm is flagging too much dissent from developing countries as "fake news" due to bias in training data, prompting corrective action. Institutionalizing this could catch problems early and create internal incentives to respect speech rights.

By implementing these recommendations, stakeholders can move toward a more balanced social media environment—one that addresses real harms (fraud, libel, incitement) vigorously, but without casting a wide chill over legitimate discourse. The result will not be a perfectly orderly discourse (which is an unrealistic and arguably undesirable goal in a free society), but rather a *robust, self-healing public sphere* that is both vibrant and anchored in truth.

These measures collectively reinforce a culture where truth is pursued through open exchange and persuasion. Importantly, they guard against the consolidation of narrative power that we have identified as the core peril in current "misinformation" regulatory trajectories. Implementing them will require political will and public support—both of which can be marshaled by highlighting what is at stake: the very freedoms that undergird our ability to live together in truth and liberty.

In making these recommendations, we recall the biblical exhortation to "speak the truth in love" (Eph. 4:15). A society that frees its members to do just that—speak truth boldly, receive truth humbly, and correct falsehood through love and reason—will be one where misinformation finds little purchase and where the light of truth can shine, unhidden, to the benefit of all.

References

Aggarwal, V. K. (2024). Platform power and regulatory capture in digital governance. *Business and Politics, 27*(2), 284-308. https://doi.org/10.1017/bap.2024.33

Altoe, F., Moreira, C., Pinto, H. S., & Jorge, J. A. (2024). Online fake news opinion spread and belief change: A systematic review. *Human Behavior and Emerging Technologies, 2024*, Article 1069670. https://doi.org/10.1155/2024/1069670

American Academy of Child and Adolescent Psychiatry (AACAP). (2023). *Policy statement*

on the impact of social media on youth mental health. https://www.aacap.org/AACAP/Policy_Statements/2023/Social_Media_Youth_Mental_Health.aspx.

Ashraf, M. M. (2024, December 16). UN global principles on information integrity: A missed opportunity. *Bot Populi*. https://botpopuli.net/un-global-principles-on-information-integrity-a-missed-opportunity/

Billingham, P. (2024). Subsidiarity, sphere sovereignty, and state sovereignty. *European Journal of Political Theory*. Advance online publication. https://doi.org/10.1177/14748851241269585

Caesmann, M., Goldzycher, J., Grigoletto, M., & Gschwent, L. (2024). Censorship to defend democracy. *CEPR VoxEU*. https://cepr.org/voxeu/columns/censorship-defend-democracy

Catholic News Service (CNS). (2021, August 20). *Faith leaders raise concerns over online censorship of religious speech*. National Catholic Reporter. https://www.ncronline.org/news/theology/faith-leaders-raise-concerns-over-online-censorship-religious-speech.

Chhabra, J., Pilkington, V., Benakovic, R., Wilson, M. J., La Sala, L., & Seidler, Z. (2025). Social Media and Youth Mental Health: Scoping Review of Platform and Policy Recommendations. *Journal of medical Internet research*, *27*, e72061. https://doi.org/10.2196/72061

Cieslikowski, J. (2022, December 22). *Half of Americans want social media companies to remove 'misinformation' and 'hate'*. Foundation for Individual Rights and Expression. https://www.thefire.org/news/half-americans-want-social-media-companies-remove-misinformation-and-hate

Douek, E. (2022). Content moderation as systems thinking. *Harvard Law Review*, 136(2), 526–606. https://harvardlawreview.org/wp-content/uploads/2022/11/136-Harv.-L.-Rev.-526.pdf

Engler, P. (2023, January 25). *God's image as the foundation for human rights*. Answers in Genesis. https://answersingenesis.org/sanctity-of-life/gods-image-as-the-foundation-for-human-rights/

Funk, A., Shahbaz, A., & Vesteinsson, K. (2023). *The repressive power of artificial intelligence.* In *Freedom on the Net 2023* (pp. 1–35). Freedom House. https://freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence

Haack, S. M. (2012). Christian explorations in the concept of human dignity. *Dignitas*, 19(3), 4–7, 10–13. https://www.cbhd.org/dignitas-articles/christian-explorations-in-the-concept-of-human-dignity

Hasinoff, A. A., & Schneider, N. (2022). From Scalability to Subsidiarity in Addressing Online Harm. *Social Media + Society*, *8*(3). https://doi.org/10.1177/20563051221126041 (Original work published 2022)

House Select Subcommittee on the Coronavirus Crisis (HSSCC). (2021, December 17). *Email correspondence between Francis Collins and Anthony Fauci regarding the Great Barrington Declaration* [FOIA release]. https://coronavirus.house.gov/sites/democrats.coronavirus.house.gov/files/2020.10.13%20FOIA-00001024.pdf

Jacobs, L. G. (2022, June 2). *Freedom of speech and regulation of fake news*. *The American Journal of Comparative Law, 70*(Supplement 1), i278–i311. https://doi.org/10.1093/ajcl/avac010

Justitia, Vanderbilt University, & Aarhus University. (2023). *The free speech recession hits home: Mapping laws and regulations affecting free speech in 22 open democracies*. Future of Free Speech Project. https://futurefreespeech.com/the-free-speech-recession-hits-home/

Kausche, K., & Weiss, M. (2025). Platform power and regulatory capture in digital governance. *Business and Politics*, *27*(2), 284–308. Doi: https://doi.org/10.1017/bap.2024.33

Kozyreva, A., Herzog, S. M., Lewandowsky, S., Hertwig, R., Lorenz-Spreen, P., Leiser, M., & Reifler, J. (2023). Resolving content moderation dilemmas between free speech and harmful misinformation. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(7), e2210666120. https://doi.org/10.1073/pnas.2210666120

Lim, G., & Bradshaw, S. (2023). *Chilling legislation: Tracking the impact of "fake news" laws on press freedom internationally*. Center for International Media Assistance. https://www.cima.ned.org/publication/chilling-legislation/

Littlejohn, B. (2021, April 16). *A theology of free speech*. The Gospel Coalition. https://www.thegospelcoalition.org/article/theology-free-speech/

McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly, 36*(2), 176-187. https://doi.org/10.1086/267990

Mitchell, D. (2020). *What does Christianity have to do with freedom of speech?* Solas CPC. https://www.solas-cpc.org/what-does-christianity-have-to-do-with-freedom-of-speech/

Office of the Surgeon General (OSG). (2021). *Confronting health misinformation: The U.S. Surgeon General's advisory on building a healthy information environment.* U.S. Department of Health & Human Services. https://www.hhs.gov/sites/default/files/surgeon-general-misinformation-advisory.pdf

Organisation for Economic Co-operation and Development (OECD). (2023, February 20). *Media literacy education system: Finland* [Case study]. https://www.oecd.org/en/publications/2023/02/mis-and-disinformation_c584ff91/media-literacy-education-system_28535133.html

Peterson-Salahuddin, C. (2024). Repairing the harm: Toward an algorithmic reparations approach to hate speech content moderation. *Big Data & Society*, *11*(2). https://doi.org/10.1177/20539517241245333 (Original work published 2024)

Pew Research Center. (2025, August 19). Widespread global public concerns about false information online. https://www.pewresearch.org/global/2025/08/19/widespread-global-public-concerns-about-false-information-online/

Pew Research Center. (2025, April 24). *Importance of press freedom, free speech, and freedom on the internet*. https://www.pewresearch.org/global/2025/04/24/importance-of-press-freedom-free-speech-and-freedom-on-the-internet/

Pozen, D. (2022, February 9). *"Truth drives out lies" and other misinformation*. Knight First Amendment Institute. https://knightcolumbia.org/blog/truth-drives-out-lies-and-other-misinformation

Robertson, K., & Williams, A. (2024). Mass media – agenda-setting theory and the propaganda model. In *Insights into Communication Studies: Understanding Foundational Theories for Media Literacy*. Open Educational Alberta. https://pressbooks.openeducationalberta.ca/insightsintocommstudies/chapter/chapter-8-mass-media/

Sangwa, S. and Mutabazi, P. (2025), Lucifer's Operating System: Exposing the Spiritual Code of Global Governance. Open Journal of Business Theology , Available at SSRN: http://dx.doi.org/10.2139/ssrn.5344142

Sangwa, S. and Mutabazi, P. (2025). Artificial Intelligence Ethics and Biblical Prophecy: A Global Christian Analysis of Algorithmic Censorship, Digital Deception, and Eschatological Risk. Available at SSRN: http://dx.doi.org/10.2139/ssrn.5283976

Sangwa, Sixbert and Mutabazi, Placide, Compliance as the New Competence: How Obedience Cultures Erode Innovation and Leadership Ethics in Global Workplaces (July 29, 2025). Available at SSRN: http://dx.doi.org/10.2139/ssrn.5371035

Sangwa S. (2025). UN Pact for the Future and Global Digital Compact: A Critical Interdisciplinary Analysis of Global Digital Governance, Digital-ID, and Central-Bank Digital Currencies. Open Journal of Stigmatized Knowledge & Suppressed Discourses (OJ-SKSD). Vol 01. Issue 01. Open Christian Press. DOI: https://doi.org/10.5281/zenodo.16983508

Santos, N. (2024, March 4). Why do we need to discuss so-called "information integrity"? *TechPolicy.Press*. https://www.techpolicy.press/why-do-we-need-to-discuss-socalled-information-integrity/

Schaefer, B. (2024, September 16). *The U.S. must oppose the U.N. pact for the future*. The Heritage Foundation. https://www.heritage.org/node/25157583/print-display

Sullum, J. (2023). *Under Government Pressure, Twitter Suppressed Truthful Speech About COVID-19*. https://reason.com/2023/01/02/under-government-pressure-twitter-suppressed-truthful-speech-about-covid-19/

U.S. Department of Health and Human Services (HHS), Office of the Surgeon General. (2023, May 23). *Social media and youth mental health: The U.S. Surgeon General's advisory*. https://www.hhs.gov/surgeongeneral/reports-and-publications/youth-mental-health/social-media/index.html

WACC Global. (2019, August 12). *From manufacturing consent to manufacturing consensus*. https://waccglobal.org/from-manufacturing-consent-to-manufacturing-consensus/

Walker, A. T. (2024, September 4). Why free speech should be a Christian priority. *WORLD*. https://wng.org/opinions/why-free-speech-should-be-a-christian-priority-1725445940

Wikipedia contributors. (2025, October 2). *Sphere sovereignty*. Wikipedia. https://en.wikipedia.org/wiki/Sphere_sovereignty

Zweig, D. (2023, January 3). *What the Twitter Files reveal about COVID-19 truths*. The Free Press. https://www.thefp.com/p/twitter-files-covid-truths